INTERFACING DATA HARNESSING, STOCHASTIC MODELING AND OPTIMIZATION FOR MAINTENANCE DECISIONS FOR RAILWAYS

A Dissertation
Presented to
The Academic Faculty

By

Mariana de Almeida Costa

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology

December 2020

INTERFACING DATA HARNESSING, STOCHASTIC MODELING AND OPTIMIZATION FOR MAINTENANCE DECISIONS FOR RAILWAYS

Thesis committee:

Dr. David Goldsman, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Nagi Gebraeel
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. António Ramos Andrade, Advisor
Department of Mechanical Engineering
IDMEC, University of Lisbon - Instituto Superior Técnico

Dr. Joel Sokol
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Brani Vidakovic
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date approved: October 12, 2020

No person ignores everything.

Nobody knows everything.

We all know something.

We all ignore something.

That is why we always learn.

– Paulo Freire

For my dear mother Enói

ACKNOWLEDGMENTS

To begin, I would like to express my sincerest appreciation to my advisor, Dr. David Goldsman, for his encouragement, kindness, and infinite patience throughout my doctoral studies. He went beyond his duties to support me in every aspect, and this has been invaluable in creating an atmosphere where I could grow academically as well as personally. I feel extremely fortunate for having the opportunity to work under his advisement.

I would also like to express my deepest gratitude to my advisor, Dr. António Ramos Andrade, who believed in my potential from day one and invited me to join one of the most enthusiastic research groups I have ever worked with. I had already been fascinated by the railway research published by Dr. Andrade even before I went to Instituto Superior Técnico. His academic guidance was definitely a crucial contribution to the completion of my research, and I am extremely grateful for having the possibility of learning so much from him. I could not have been advised by a better pair of advisors.

I also want to thank the rest of the committee, Dr. Brani Vidakovic, Dr. Nagi Gebraeel and Dr. Joel Sokol for reading through this thesis, and for all the feedback they have given me during my Ph.D. studies.

I would like to specifically thank Dr. Joel Sokol and Dr. David Goldsman for allowing me the opportunity to work as their Teaching Assistant in the Master's Program in Analytics at Georgia Tech. This experience was a turning point in my career, showing me not only that teaching is the best way to learn, but also that I am truly passionate about it. I became more confident in my potential and started to think about a career in academia because of my experience as a TA for ISyE6501 and ISyE6644.

I would like to give special thanks to all my research colleagues and friends at Instituto Superior Técnico. Their insightful comments, questions and support were a significant contribution to the completion and quality improvement of my research project. In

particular, I would like to thank Joaquim Braga, João Costa, João Pagaimo and Luís Sancho for patiently showing me their research work at IST, explaining the databases and the challenges. You, and all the other colleagues, have made my workdays in Lisbon very enjoyable and I will always carry these good memories with me.

I also need to acknowledge VLI Logística, MRS Logística and Fertagus maintenance teams, in the name of Hugo Ribeiro (MRS Logística), Raphael Damasceno and Gustavo Castro (VLI), João Grossinho and João Duarte (Fertagus). The projects we worked on together were crucial for the development of this research and for my understanding of the opportunities related to track and wheelset maintenance. I appreciate their patience and invaluable advice.

In addition, I would like to acknowledge the Georgia Institute of Technology and, specifically, the School of Industrial and System's Engineering. My deepest appreciation to Dr. Santanu Dey, Dr. Alan Erera, Ms. Amanda Ford, Ms. Jennifer Wooley, Dr. Dawn Strickland, Dr. Harry Sharp and all the faculty and staff members that make ISyE the outstanding school it is. I feel extremely proud of being part of this community of incredible people, and I cannot tell how much I value all the experiences and knowledge that I received during my stay here.

My deepest thanks to the Brazilian National Council for Scientific and Technological Development (CNPq) and the "Science Without Borders" program, administered by LASPAU, for providing me with financial support during part of my Ph.D. studies (Science Without Borders Program - LASPAU - GDE - EUA - grant 203130/2015-4).

My sincere appreciation to all my friends in Atlanta and ISyE. You have become my family in the United States, and I really appreciate your encouragement and friendship during my Ph.D. years. I will always remember the great times we shared together and I really appreciate you being there along the way, I would not have made it without you all.

Last but not least, I would like to thank my parents, my brother and all my relatives and friends in Brazil. Moving out to Atlanta to pursue the Ph.D. dream would not have been

possible without you. I would like to give special thanks to my mother Enói. I have no words to describe everything she has done for me throughout my life. In her twenties, she would work all day and take night classes to complete high school. She could not afford to go to college, but she literally dedicated her life to make sure my brother and I had access to quality education. She dreamed of me becoming a doctor probably more than I did and she believed in me even when I did not. This thesis is dedicated to her, for all her support, the infinite hours on the phone telling me words of encouragement and her unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The increasing demand for cost-effective and transparent solutions for the improvement of the maintenance decision-making process in railways fuels the development of more sophisticated and flexible models, which largely exploit the use of data analytics and optimization tools. At the same time, recent advancements in technologies for railway condition monitoring and the availability of massive amounts of data allow for more accurate and reliable fault detection. One obstacle, however, is how to deal with the data provided by the monitoring equipment as well as the choice of suitable methods to translate the data into useful information for maintenance scheduling and prioritization. In light of this, three main stages of the maintenance decision-making process can be identified: i) data acquisition, ii) modeling approach and, iii) implementation of the policy. Deciding on which parameter(s) represent the real condition of the asset and accurately measuring them, guaranteeing appropriate instrument and good measurement precision concerns data acquisition (step i)). Next, step ii) implies the choice of a comprehensive model that can tackle all the constraints and uncertainties associated with the deteriorating system, while providing solutions (in terms of a maintenance policy) in a reasonable amount of time. Finally, step iii) concerns the ease of implementation of the new maintenance policy, guaranteeing its practical applicability within the context of the train operating company under study.

This dissertation aims to provide contributions to these three aspects in terms of railway track and wheelset maintenance. For both deteriorating systems, the choice of an appropriate maintenance policy should balance the trade-off between maintenance costs

and costs resulting from the poor-maintained asset, including those arising from potential safety hazards. This is discussed in the context of the three main stages mentioned above.

The dissertation is structured in five chapters. Chapter 1 provides the introduction, as well as a brief overview of each of the topics and results presented in the subsequent chapters. Then, chapters 2 and 3 focus on wheelset maintenance and chapters 4 and 5 focus on railway track maintenance.

In chapter 2, the optimization of railway wheelset maintenance policy is discussed. This policy is developed based on a data-driven model encompassing estimation of wear rates and further application of a Markov Decision Process (MDP) approach to represent possible discretized wheel states, where the problem of maintenance planning is tackled from the perspective of immediate action cost-optimization. A bidimensional framework considering discrete intervals of wheel diameter along with a quantitative variable (kilometers since last turning/renewal) is used to represent the possible wheel states. In addition, the probability of a defect interfering with the wheel maintenance schedule is modeled by contemplating survival curves derived from a Cox Proportional-Hazards model. As a secondary goal, a comparison of the optimized policy with another wheel's reprofiling policy that is also "easy to implement" is provided.

In chapter 3, an investigation around the uncertainty of wheelset inspection data is made. Previous research has highlighted the relevance of this topic in the decision-making process surrounding wheelset maintenance actions. In light of this, the investigation aimed to assess the agreement between data acquired from three different inspection devices, namely: i) manual (gauge device), ii) a laser device and iii) an under-floor wheel lathe. Three main wheelset parameters (flange thickness (Ft), flange height (Fh) and flange slope (qR)) are compared using a Linear Mixed Model (LMM) approach under several real-world limitations, such as those imposed by serially correlated, unbalanced and unequally replicated data. Findings supported the use of LMM, showing its ability to capture and account for the differences among the various groups and highlighting statistical significant

performances of the inspection devices.

In the context of the railway track, chapter 4 presents a spatiotemporal approach for the modeling and prediction of track geometry faults. Spatial-time data from a train operating company is considered through a 5-year inspection database. The track twist, defined as the amount by which the difference in elevation of rails increases or decreases in a given length of the track, is used as the main track quality parameter. The spatiotemporal approach considered two Kriging models with a Gaussian correlation function to study a strategic portion of a track used in heavy-haul transport. A CUSUM (Cumulative Sum) control chart approach is then applied to identify out-of-control track sections and a Logistic Regression model is used to get estimates of the probabilities of future out-of-control points based on the adopted thresholds. Finally, a simple MDP model based on out-of-control points is proposed to compare different maintenance policies aimed at cost minimization for different thresholds of twist standard deviation for different track sections grouping strategies.

Lastly, chapter 5 explores the use of Wavelet Analysis (WA) in the statistical modeling of railway track irregularities, namely (1) longitudinal level, (2) alignment, (3) cross-level, (4) gauge. WA is used to study and reconstruct the four different track geometry irregularity signals. This investigation aimed at finding wavelets that can appropriately describe each track irregularity signal studied, and investigating whether the presence of some high amplitude wavelet coefficients in certain frequencies can be associated with higher vertical or lateral forces in the wheel-rail contact. The last step is accomplished by reconstructing the different irregularities signals using wavelets coefficients in various decomposition levels and studying their impact on Nadal's safety criterion Y/Q (a critical quantity for derailment safety assessments) through vehicle dynamics simulations. The presence of certain wavelets at different decomposition levels allows identifying wavelets that are more prejudicial in terms of the safety criterion.

CHAPTER 1
INTRODUCTION

The worldwide increase in the demand for goods and transportation services boosts the need for railway cargo transport. At the same time, the increased flow of passengers in transportation networks enhances the challenges in the operation of passenger trains. Therefore, railways play a pivotal role in transportation systems worldwide. Effective maintenance policies are crucial to guarantee the reliability and profitability of such systems.

Existing standards, such as EN 13848 [1] and UIC 714 [2], provide guidelines for maintenance activities of railways, promoting interoperability and harmonization across different companies operating in different countries. In this sense, these standards are often conservative, as they do not take into consideration the particularities of each train operating company, such as environmental conditions, track geometry and design (e.g.: materials, curvature, track slope, the existence of substructure shortcomings).

On the other hand, even strict compliance with the conservative guidelines from standards does not guarantee the non-occurrence of failures. Therefore, in addition to following railway standards, many train operating companies also have specialized teams dedicated to i) reduce the number of unnecessary, early maintenances, ii) increase the useful lifetime of the assets, iii) minimize the number of unexpected failures, and iv) diminish the impact of maintenance on operations. Regarding iii), in terms of railway track, unexpected failures can cause, for example, from a simple train speed restriction, to potentially a total infeasibility of circulation, under risk of a safety hazard. The other three aspects are intrinsically related to improving efficiency by increasing availability and reducing costs while ensuring the quality and safety of services in a very competitive transportation environment.

In this dissertation, some contributions are made in terms of frameworks that consider the availability of condition monitoring and/or inspection data to accurately characterize the degrading railway system and propose maintenance policies with significant improvements in cost and reliability. The policies consider both the economic and stochastic dependence between the various maintenance decisions, thus providing the Infrastructure Managers (IMs) relevant information on how to choose the most appropriate policy based on a better understanding of the degradation patterns and their relationship with costs.

In the next subsections, short introductions to the railway track and the railway wheelset are provided, as well as a quick walkthrough of the research problems and main contributions.

1.1 The Railway Wheelset

A railway wheelset consists of two wheels linked by a rigid axle and it allows the motion of the vehicle when the wheels roll over the rails, as shown in Figure 1.1.

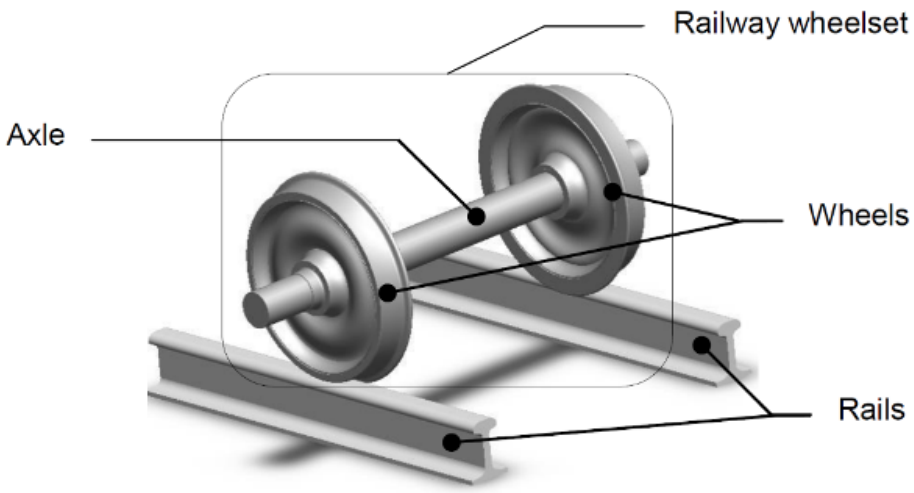

Figure 1.1: Schematic representation of a railway wheelset

Railway wheelsets are important components of trains which are affected by wear. Such high reliable products are designated to operate without failure for a long time,

hence estimating their failure-time distribution is difficult, as only a few units will fail or significantly degrade in a test of practical length at normal use conditions [3].

In terms of monitoring of railway wheels, Alemi et al. [4] point out the differences in terms of data acquisition. In-service and in-workshop inspection, wayside and onboard measurement and diagnostic and prognostic approaches are some examples. Maintaining the wheel profile is becoming increasingly important, as wear has serious influences for the safety and comfort of train operation, not to mention the life span of the rail. In addition, excessive re-profiling of the wheels can result in significant extra costs incurred for the train operating company, affecting its competitiveness [5]. Therefore, optimizing the wheelset maintenance strategy is an attractive research topic.

1.2 The Railway Track

The track is a part of the infrastructure of the railway system, which is comprised of four subsystems: track, signaling, electrical and telecom [6]. Track itself includes a superstructure and a substructure. The superstructure is comprised of rails, sleepers and the fastening system and the substructure is formed by the ballast, the subballast and subsoil [7]. Many geometric parameters were created and they are widely used to measure the condition of the track.

Following Soleimanmeigouni et al. [8], track geometry parameters can be divided into five classes: (1) longitudinal level, (2) alignment, (3) cross-level, (4) gauge, and (5) twist. The longitudinal level is the track geometry of track centreline projected onto a longitudinal vertical plane. Alignment is the track geometry of track centreline projected onto a longitudinal horizontal plane. Cant (cross-level) is the difference in height of the adjacent running tables computed from the angle between the running surface and a horizontal reference plane. Gauge is the distance between the gauge profiles of two adjacent rails at a given location below the running surface. The twist is the algebraic difference between two cross-levels taken at a defined distance apart, usually expressed

as a gradient between the two points of measurement [8]. Figure 1.2 shows the graphical representation of a railway track with its gauge and cant.

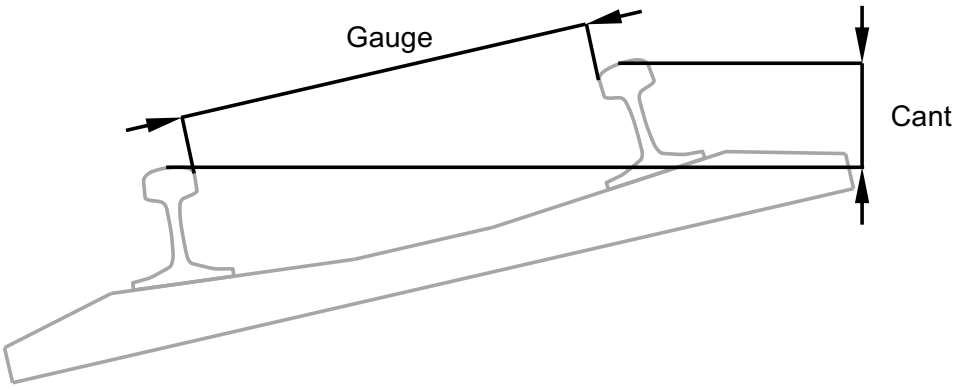

Figure 1.2: Schematic representation of track cant and gauge

Track geometry deteriorates under the influence of dynamic track loads. These loads cause stresses and elastic displacement and, depending on the total stress level, also permanent deformations [9]. As observed by Mohammadzadeh et al. [10], the random nature of these geometric irregularities, wear on the rail profile, variations in track stiffness, and track structural issues in addition to deterioration of systems used in the railway fleet are sources of the stochastic nature of track–rail interaction.

The main processes of track deterioration are wear, fatigue and settlement [11]. As track deteriorates over time, inspections are essential to gather knowledge about its current condition and monitor various parameters, minimally guaranteeing that they lie within the tolerance limits established in the standards, such as UIC 518 [12] and EN 13848 [1]. Periodic inspections through special vehicles that run the track are widely used as an instrument to measure the different track parameters and identify possible defects, by means of signal digital processing techniques [13, 14, 15, 16]. Often, these signal processing techniques include the utilization of filters that focus on particular wavelength band and eliminate wavelengths outside that band [14]. Then, the track is divided into sections (and/or segments), i.e. lengths of track where traffic, ballast and sleeper types are assumed constant [17], and standard deviations of the filtered signals are computed for each

track section. UIC 518 [12] and EN 13848 [1] are examples of standards that include alert limits for geometric parameters using this type of methodology.

1.3 Overview of Research Problems and Main Contributions

The first part of this dissertation, chapters 2 and 3, focuses on two problems related to railway wheelset maintenance. Wheelsets are important components of trains which are affected by wear and absorb a significant part of the maintenance budget of any train operating company.

The two case studies presented in chapters 2 and 3 use a 17-year interval inspection database, ranging from January 2001 up to July 2019. Wheel measurement data was acquired from a Portuguese railway company that transports passengers in a single line, which extends 54 kilometers and serves 14 stations.

Chapter 2 is based on the paper [18]. Inspection data is used to model both degradation and likelihood of damage, in terms of estimation of wheel wear rates and survival curves, respectively. These are further incorporated into a Markov Decision Process (MDP) model. An MDP refers to a general mathematical framework for sequential decision making under uncertainty when the system can be successfully represented by states. A bidimensional framework considering discrete intervals of wheel diameter along with a quantitative variable (kilometers since last turning/renewal) is used to represent the possible wheel states, while the probability of a defect interfering with the wheel maintenance schedule is modeled by contemplating survival curves derived from a Cox Proportional-Hazards Model (CPHM). The inclusion of damage modeling through CPHM is one of the main contributions of the work described in chapter 2. A thorough review of MDPs and CHPM, as well as background, is also provided in the chapter.

The output of the MDP model comes in the form of a map of decisions, in terms of actions to be taken for each discretization of kilometers since last turning/renewal and wheel diameter, where two distinguished wheelset policy paths can be identified. Optimal

results in terms of minimal cost policy are discussed in the context of the MDP, but a more realistic and easy-to-implement policy fixing one of the parameters is compared to the optimal policy, which is another main contribution of the work developed. Results have shown that, in practice, train operating companies might benefit from using the easy-to-implement policy, which has an associated long-run average cost only about 1% higher than the one suggested by the optimal decision map.

Chapter 3 is based on paper [19] and it addresses the uncertainty around inspections in terms of wheelset measurements. Data acquired from the passenger train operating company on wheel profile for three different inspection devices is again considered, namely: i) manual (gauge device), ii) a laser device and iii) an under-floor wheel lathe. Availability of such data motivates the investigation of differences in precision among the devices. Three main wheelset parameters (Flange Thickness (Ft), Flange Height (Fh) and Flange Slope (qR)) are compared using a Linear Mixed Model (LMM) approach. Previous research, as shown in chapter 3, has highlighted the importance of reliable data in the decision-making process surrounding wheelset maintenance actions. Assessing agreement using inspection data from a system in operation under different conditions than the commonly adopted ones for controlled studies, such as repeated measurements and balanced data, is very rarely done in the scientific literature and has its limitations. On top of that, when the true measurement values are unknown, a lot of methods are of impractical use. Although previous works have considered LMMs and inspection data, they have not explicitly dealt with the assessment of the agreement, which is the main contribution of chapter 3.

The problem of assessing agreement is addressed by assuming that an LMM can explain the association between the response and the fixed predictors. Then, controlling for the variation of different groups by adding random intercepts, allows for comparison between measurement types by performing residual analysis, which was the methodology adopted. The mixed-model framework provides a systematic way to account for the between- and

within-subject variability and for the serial correlation among measurements, making it possible to compare the spread in the distribution of measurements coming from the different devices.

Chapter 3 also presents some interesting and relevant results encompassing railway inspection data and the use of mixed-model framework. First, a comparison of linear models and the mixed-models approaches, indicating the superiority of the LMMs. The next relevant result is related to the behavior of the residuals assuming the underlying models were correct. In this case, data did not show evidence supporting small differences in the variances for the measurement devices. These differences were only statistically significant for the flange height parameter and for the comparison in the pair manual and wheel lathe measurements for the flange slope parameter. The statistical test for the flange thickness parameter did not reveal significant differences in terms of variances.

In a nutshell, the analysis in chapter 3 shows that some of the mechanisms related to the inspection activities inside maintenance shops may leverage the impact of the device performance. In addition, the flange thickness parameter seems to be less prone to differences in precision when compared to the other parameters. Therefore, it may be beneficial for IMs to use the flange thickness as a more robust parameter for the actual wheel condition.

The second part of the dissertation, chapters 4 and 5, presents two studies encompassing railway track. More specifically, both problems deal with track geometry deterioration. A deeper understanding of the features associated with track deterioration and the impacts of maintenance in the reliability of the system plays a fundamental role in a better maintenance decision-making process. This can be achieved by creating maintenance frameworks that make a more accurate characterization of the remaining lifetime distribution of the track geometry components and can successfully translate this into the whole system reliability considering its dynamic and correlated nature, while also providing the IMs information and flexibility for defining and changing the maintenance window according to schedules

that exploit potential economic savings of maintenance decisions for different grouping strategies.

Chapter 4 is based on paper [20]. Spatial-time data from a railroad company operating in Brazil is considered through a 5-year inspection database. The signal data is comprised by measurements of track geometry parameters, collected at every foot. A total of nine inspections, spaced six months apart from each other, are considered.

Track geometry deteriorates through time, and a good maintenance strategy should keep the track's condition at an acceptable level; this can be achieved by monitoring various important parameters and ensuring they are under control. One such parameter used to assess track geometry condition is the track twist, defined as the amount by which the difference in elevation of rails increases or decreases in a given length of the track. The spatiotemporal approach in chapter 4 considers two Kriging models, namely ordinary Kriging (OK) and limit Kriging (LK), with a Gaussian correlation function to study a strategic portion of a track used in heavy-haul transport.

The nature of the collected data makes it spatial and time autocorrelated. Intuitively, if track is degraded at some point in time and no event (such as maintenance activity) dramatically changes this situation, it is likely that the next measurement will point out a degraded track, which is a motivation to use a spatiotemporal model. In particular, the OK model was chosen due to the fact that the twist values tend to revert to a mean, as a negative twist is usually followed by a positive twist. An alternative method with LK was also proposed for contrasting purposes, especially in the points where the measured signal peaks, indicating possible future failure points. The two Kriging models were trained with 4-year data (a total of eight inspections) and applied to find one-step-ahead extrapolation in the time domain, and the comparison of performance, in terms of test data error of the two approaches, is one of the main contributions of this work.

In possession of the predictions, a Statistical Process Control (SPC) technique known as CUSUM (Cumulative Sum) control chart was applied to identify out-of-control track

sections. Then, a Logistic Regression model was used to get estimates of the probabilities of future out-of-control points based on the adopted thresholds. Finally, a simple MDP model based on out-of-control points was proposed to compare different maintenance policies aimed at cost minimization for different thresholds of twist standard deviation. Another contribution of chapter 4 was to show the cost percentual differences arising from different grouping strategies of track sections, taking into account the impact on availability given that maintenance is performed in a single line track, which implies traffic blockage whenever maintenance is needed.

To conclude the dissertation, chapter 5 explores the use of Wavelet Analysis (WA) in the statistical modeling of railway track irregularities, namely (1) longitudinal level, (2) alignment, (3) cross-level, (4) gauge. Data was acquired from a Portuguese train operating company, encompassing a 2048 m line, consisting of UIC 60 kg/m rail profiles laid out in Iberian gauge (1.668 m). Chapter 5 is based on paper [21].

The investigation presented in chapter 5 is a contribution to a recent debate topic within track maintenance which refers to the use of predefined standard alert limits, such as those found in EN 13848-5 [1], and presents an alternative assessment method based on wavelets. The methodology adopted combines WA, which is used to study and reconstruct different track geometry irregularity signals, and vehicle dynamics simulations to identify track faults by using the traditional Nadal's safety criterion for derailment, known as Y/Q. The goal was to find evidence of wavelets and respective scales which could be associated to increased Y/Q values, especially those surpassing the established limit of 0.8 for the safety of trains operations, even when the original track geometric signals measured at the extrapolation points had not surpassed the immediate action limit (IAL) suggested in the conventional track maintenance standards, in this case, the European Standard EN 13848-5 [1]. The working hypothesis is that the various defects have a range of varying wavelength and frequency content, and their observance in some specific "shapes" in the original signal can be an indicator of track faults that are potentially more penalizing in terms of risk of

derailment.

Two main contributions of the experimental investigation conducted in chapter 5 are: defects occurring in some scales/frequencies are potentially more harmful from the standpoint of Y/Q than other defects in other scales/frequencies, and the similarity of some particular geometric defect with the ”shape” of some wavelet may be used to identify points where to target maintenance actions. These findings raise questions on the appropriateness of applying the same track maintenance standards to all train operating companies and track irregularities based on point-wise comparisons with alert limits.

CHAPTER 2
AN OPTIMIZED MAINTENANCE FRAMEWORK FOR RAILWAY WHEELSETS

2.1 Introduction

The increased flow of passengers in transportation networks enhances the challenges in the operation of passenger train. Minimizing disruption of train schedules while guaranteeing an adequate maintenance plan is mandatory to keep the railway network services competitive. In this context, obtaining cost-effective thresholds becomes an important goal within the maintenance strategy [16].

From an economic point of view, railway maintenance is a relevant topic, as high capital costs are involved. Wheels deteriorate through time and prolonging their useful life translates directly into savings, by avoiding unnecessary downtime and costs of renewal and re-profiling. In this sense, regular inspections play a fundamental role, as monitoring various important parameters and ensuring they are under control allows re-profiling and replacement actions to be undertaken very closely to cost-effective targets. Nevertheless, commonly adopted international standards providing guidelines for secure interoperability of wheelsets may be too conservative and fail to approximate the actual wear rates, since the latter is influenced by a combination of factors that uniquely exist in the context of the train operating company, such as environmental conditions, track geometry and design (e.g.: materials, curvature, track slope, existence of substructure shortcomings), among others.

Routine inspection is a common practice among train operating companies, usually performed based on time or distance intervals. For the wheels, geometrical parameters such as diameter (D), flange thickness (Ft), flange height (Fh) and flange slope (qR) are

collected and compared to the limits established in the standards [22]. Figure 2.1 provides a schematic representation of a railway wheel profile and the four parameters which are measured relative to three fixed measurements (a, b, c) and to a tread datum (T) position (70 mm measured from the flange back).

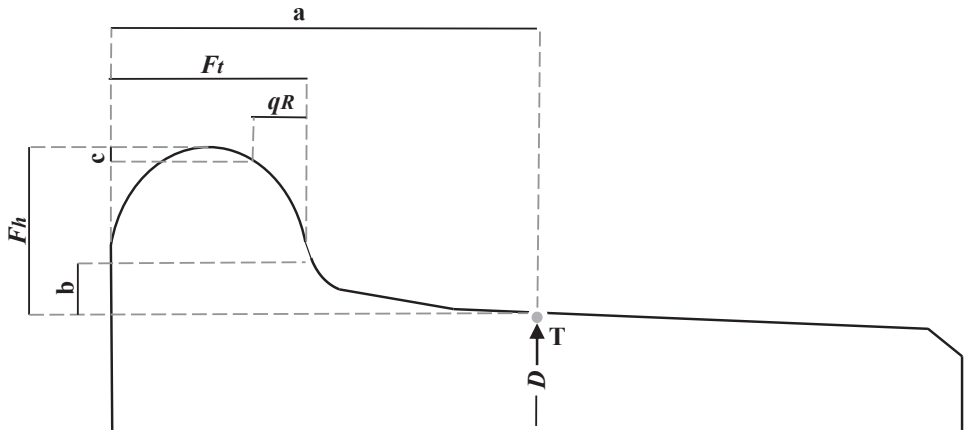

Figure 2.1: Diameter (D), flange height (Fh), flange thickness (Ft), and flange slope (qR)

2.2 Related Work

In a recent review of railway transportation, Ghofrani et al. [23] reported an increased interest in the field for studies focused on a higher level of safety with reduced maintenance costs, as well as preventive and conditional maintenance tasks. More specifically, they show that recent literature on wheels mostly considers either a preventive or condition-based maintenance approach. Growing interest on wheel wear is justified based on an increase of problems originating from accelerated wear, due to heavy haul, growing traffic, and increased frequency of operation [24].

Many different models for wheel profile deterioration and prediction are available in the literature. Some recent approaches are based on mechanistic models in conjunction with vehicle dynamics [24, 25, 26], some others include controlled experiments to estimate reliability and remaining useful life [27, 28]. Statistical approaches for the estimation of wheel wear are also presented in some works.

Shebani and Iwnicki [29] use Nonlinear Autoregressive models with exogenous input neural network (NARXNN) to predict wheel and rail wear under different contact conditions, based on experimental data obtained in laboratory tests, and conclude that artificial neural networks can be used efficiently as a predictor of wheel and rail wear, with an accuracy of the model surpassing 80%. Cremona et al. [30] point out that wear coefficients are currently derived through laboratory tests or extensive calibrations based on geometrical comparisons between simulated and measured wheel profiles, however, the resulting coefficients are not accurate as the experiments are limited in the sense of reproducing the actual wheel and track conditions. Therefore, providing a measure of the uncertainty around the estimated coefficients is advisable. They present a novel approach to estimate the wheel wear by exploiting the correlation in measurements with similar contact pressure and sliding speed through a universal kriging approach, enabling uncertainty analysis on the wear coefficient and construction of wear prediction intervals. The results showed that considering the uncertainty around the wear coefficient led to much better predictions than traditional models. In the latter, not allowing the wear coefficient to change, or just using an average of it, resulted in underestimated and overestimated (respectively) wear volume. Finally, some other studies focus on models for the prediction of specific defects, such as flats [31] and rolling contact fatigue [32], or study/prediction of multiple failure modes [33, 34, 35].

For wheels developing appreciable wear takes a considerable amount of time, making experimentation on the actual track expensive and inconvenient, as it can interfere with railway network operations [24]. At the same time, laboratory experiments fail to replicate all possible wear mechanisms encountered in real railway operations, such as dynamic variations in load and position and a wide range of geometry variations [29]. Nevertheless, wheel wear rates are very rarely characterized by using inspection data in a real-world case study.

Markov Decision Processes (MDPs) have been extensively used in the literature for

determining and evaluating maintenance policies for deteriorating systems. Faddoul et al. [36] classify existing approaches into (i) deterministic MDPs, where perfect inspections and perfect maintenance actions are implicitly assumed; (ii) probabilistic MDPs, where perfect inspections and uncertain maintenance actions are implicitly assumed; or (iii) Partially Observed Markov Decision Process (POMDP), where the state of the system at each decision epoch is not precisely known and it deals with uncertainty around maintenance actions. Some recent applications of MDPs to maintenance problems are discussed next.

Sun et al. [37] consider an MDP model to study optimal inspection and replacement condition-based maintenance for a multi-unit system, where the degradation of each component is assumed to follow a Wiener process. More specifically, maintenance policy for K-out-of-N systems is optimized in terms of incurring costs when three different actions can be undertaken at each time epoch (inspection): 1) preventive maintenance is applied to all N components; 2) preventive maintenance is applied to the component that has a higher degradation rate, or; 3) nothing is done until next inspection. Also dealing with multi-unit systems, Faddoul et al. [36] propose an MDP model for the maintenance costs of reliability constrained series systems. Since interdependencies arise naturally when components of a system are competing for the same resources, such as limited availability of budget, maintenance parts and/or spare parts, optimal maintenance decision of one component will depend on the state of that component and on the state of the other components. The presented model includes target reliability levels and, as such, the MDP-based methodology minimizes the maintenance costs in a reliability constrained series system context, where Lagrangian relaxation techniques are used to separate the initial combinatorial problem into smaller sub-problems, allowing components to be heterogeneous.

Nguyen et al. [38] consider the problem of adjusting condition monitoring quality characterized by the observation noises on the system degradation level returned by an inspection. The optimization problem is formulated based on a generalization of MDP,

the Partially Observed Markov Decision Process (POMDP), and the authors conclude that the dynamic policy contemplating adjustment of the quality of inspections outperforms the classical fixed-quality policies. Ivy and Nembhard [39] also take advantage of POMDP, for the case where maintenance is imperfect, i.e., maintenance actions whose effect depends on the current state of the system. The model reflects the management decision with partial information, where an action is taken based on current information or delayed until more information is acquired.

In the context of railways, Gerum et al. [40] present an application of MDP to determine optimal maintenance policies for rail and track geometry defects. Using a finite discrete state space, they define each state as a composition of the current level of deterioration and the load for the next time period. Optimal policies are provided for scenarios where crew limitations exist or not.

Only a few contributions exist in the literature encompassing the study of the long-term behavior of the wheelset and identifying optimal maintenance policies including both degradation and recovery modelling. Some exceptions are briefly discussed below.

Jiang et al. [41] present a bidimensional wear model considering the wheel diameter and flange thickness, where the maintenance policy allows the threshold for re-profiling and the recovery value to be optimized conditioned on the wear state of the wheel. A Semi-Markov Decision Process (SMDP) framework is proposed with the objective of minimizing the long-run expected average maintenance cost per unit time while considering the effects of imperfect maintenance. In an extension of this work, Mingcheng et al. [42] claim that wheels are subject to external shocks, such as tread peeling and tread flat, that need to be removed immediately by cutting material off the wheel. In an SMDP framework, the same bidimensional model considering the wheel diameter and flange thickness is proposed, although it accounts for shock-based failures in the wheel re-profiling policy. Braga and Andrade [43] explore the wheel re-profiling problem through an MDP approach by considering a model that uses the wheel diameter along with a quantitative variable

(mileage since last turning) and also incorporates the transition to states with damage. As a result, maintenance guidelines based on mileage since turning are compared across discretized values of wheel diameter.

The last two studies above incorporate the transitions to damaged states. In particular, Mingcheng et al. [42] assume that shock-based failures follow a homogeneous Poisson process, where the magnitudes of shocks are independent. The parametric approach results in exponential failure times. In the study by Braga and Andrade [43], the transition probabilities to damaged states are obtained through an empirical logit model. Hence, the first work relies on a parametric approach with resulting exponential failure times, whereas the second one uses an empirical approach that completely relies on the data acquired, which could potentially lead to underestimated probabilities due to disparities found in the ratio of damaged and non-damaged wheels. The formation of defects poses an additional challenge to the modeling of the wheel lifecycle, as it is a very complex phenomenon often associated with many factors, such as contact stress, wheel and rail materials, lubricant and so on [32]. Damages affect the operational safety of trains and shorten significantly the wheelset useful life, as correcting it may require a much bigger diameter loss in comparison to a simple parameter threshold violation. At the same time, wheel damages cause an increase in attrition and damage to the railway infrastructure, also shortening its life span and leading to a reduced lifetime and availability of rolling stock [34].

2.2.1 Main Contributions

This study extends some of the ideas presented in the above subsection. An MDP approach is applied to railway wheelset maintenance, using data of wheelset maintenance activities from a Portuguese train operating company. A bidimensional framework is used to represent the possible wheel states and optimal results in terms of minimal cost policy are discussed in the context of the MDP. As main contributions, it should be highlighted: the modeling of the probability of a defect interfering with the maintenance of wheels

schedule by contemplating survival curves coming from a Cox Proportional-Hazards model and the comparison of an easy-to-implement policy to the optimal one obtained from the bidimensional MDP model.

2.3 Preliminares

This section is devoted to preliminaries. In particular it makes a review on the relevant concepts essential for the development of the proposed work.

2.3.1 Markov Decision Process

A stochastic process $\{X_n, n = 0, 1, 2, \cdots\}$ that takes on a finite number of possible values and for which the conditional distribution of any future state X_{n+1}, given the past states $X_0, X_1, \cdots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state is called a Markov Chain [44]. An MDP is a sequential decision process for which the decisions produce a sequence of Markov Chains with rewards.

Under each one of the maintenance policies, a sequence of states$\{X_n, n = 0, 1, 2, \cdots, N\}$ will constitute a Markov Chain with transition probabilities $p_{i,j}$. For every state i in the set of possible states $s \in \{s_0, s_1, s_2, \cdots, s_N\}$, an action is chosen from the set of possible actions $a \in \{a_0, a_1, a_2, \cdots, a_M\}$ (assumed finite), and for each action a set of state and action-dependent rewards/costs $q(i, a)$ and a set of state and action-dependent transition probabilities $p(j|i, a)$ are established. Hence, if the process is in state i at time n and an action a is chosen, the next state of the system, j, is determined according to $p(j|i, a)$ or, simply, $p_{i,j}(a)$. By letting X_n denote the state of the process at time n and a the action chosen at time n:

$$\Pr(X_{n+1} = j|X_0, a_0, X_1, a_1, \cdots, X_n = i, a_n = a) = \Pr(X_{n+1} = j|X_n = i, a_n = a) = p_{i,j}(a) \tag{2.1}$$

Thus, the transition probabilities are functions only of the present state and the

subsequent actions, in accordance with the Markov Property. Since probabilities are nonnegative and since the process must make a transition into some state, it follows that for a given action $a \in \{a_0, a_1, a_2, \cdots, a_M\}$, the transition probabilities from a state to any other state (including the current state) must sum up to one, i.e.:

$$\sum_{j=0}^{s_N} p_{i,j}(a) = 1, \qquad i = 0, 1, \cdots, s_N \tag{2.2}$$

At any time period n and for a given action $a \in \{a_0, a_1, a_2, \cdots, a_M\}$, the probability values can be represented in the form of a transition matrix $\boldsymbol{P}(a)$, also called Markov Transition Matrix (MTM), as follows:

$$\boldsymbol{P}(a) \equiv [p_{i,j}(a)] = \begin{bmatrix} p_{0,0}(a) & p_{0,1}(a) & p_{0,2}(a) & \cdots & p_{0,s_N}(a) \\ p_{1,0}(a) & p_{1,1}(a) & p_{1,2}(a) & \cdots & p_{1,s_N}(a) \\ p_{2,0}(a) & p_{2,1}(a) & p_{2,2}(a) & \cdots & p_{2,s_N}(a) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{s_N,0}(a) & p_{s_N,1}(a) & p_{s_N,2}(a) & \cdots & p_{s_N,s_N}(a) \end{bmatrix} \tag{2.3}$$

The matrix above brings the one-step transition probabilities. One of the major advantages of using Markov models is the ease of computation of the probabilities of visiting future states [45]. More specifically, the n-step transition probability matrix for a given action $a \in \{a_0, a_1, a_2, \cdots, a_M\}$ can be easily computed by multiplying the matrix $\boldsymbol{P}(a)$ by itself n times, this is, $\boldsymbol{P}^n(a)$. Therefore, assuming the transition probability matrix is stationary, computation of future states probabilities in a Markov process only requires

the knowledge of $\boldsymbol{P}(a)$ and the row vector of initial states probabilities $\boldsymbol{X}_0$, as follows:

$$
\begin{aligned}
&\boldsymbol{X}_0 \\
&\boldsymbol{X}_1 \;=\; \boldsymbol{X}_0 \cdot \boldsymbol{P}^{(0)}(a) \;=\; \boldsymbol{X}_0 \cdot \boldsymbol{P}(a) \\
&\boldsymbol{X}_2 \;=\; \boldsymbol{X}_1 \cdot \boldsymbol{P}^{(1)}(a) \;=\; (\boldsymbol{X}_0 \cdot \boldsymbol{P}(a))\boldsymbol{P}(a) \;=\; \boldsymbol{X}_0 \cdot (\boldsymbol{P}(a)\boldsymbol{P}(a)) \;=\; \boldsymbol{X}_0 \cdot \boldsymbol{P}^2(a) \\
&\boldsymbol{X}_3 \;=\; \boldsymbol{X}_2 \cdot \boldsymbol{P}^{(2)}(a) \;=\; (\boldsymbol{X}_0 \cdot \boldsymbol{P}^2(a))\boldsymbol{P}(a) \;=\; \boldsymbol{X}_0 \cdot (\boldsymbol{P}^2(a)\boldsymbol{P}(a)) \;=\; \boldsymbol{X}_0 \cdot \boldsymbol{P}^3(a) \\
&\cdots \\
&\boldsymbol{X}_n = \boldsymbol{X}_0 \cdot \boldsymbol{P}^n(a)
\end{aligned}
\tag{2.4}
$$

In an MDP framework, it is desired to find the set of actions that minimize (maximize) the sum of all costs (rewards) or an average cost (reward), over the set of solutions that are feasible for each state, i.e., it is desired to find an optimized policy. Each time the system visits state i at epoch n, a cost is incurred (or a reward is earned). The associated reward/cost vector, $\boldsymbol{q}$, is usually assumed to be stationary and the expected value for the total reward/cost after n epochs or time periods $\left(R_{(n)}\right)$, can then be easily computed as $E\left[R_{(n)}\right] \;=\; \boldsymbol{X}_0\boldsymbol{P}^n\boldsymbol{q} \;=\; \boldsymbol{X}_n\boldsymbol{q}$. Then, the optimization problem to determine the set of actions that minimize the discounted cost incurred over an infinite horizon solves [46]:

$$
\nu(i) = \min_{a\in\{a_0,a_1,a_2,\cdots,a_M\}} \{c(i,a) + \sum_{j\in S} \gamma \cdot \nu(j) \cdot p(j|i,a)\}, \qquad \forall i \in S \tag{2.5}
$$

where $\nu(i)$ denotes the optimum expected discounted costs incurred from epoch n onwards for a current state i, $c(i,a)$ denotes the expected cost incurred for state i under action a, $\gamma \in [0,1]$ represents the discount factor and $p(j|i,a)$ denotes the probability of making a transition to a new state j given the current state i and the action, a, taken in the current state. For further details on MDPs, the reader may refer to Puterman [46].

2.3.2 Cox Proportional-Hazards Model

The CPHM was originally proposed by Cox [47], who considered age-specific failure rates (censored failure times) to propose a regression model to investigate the association between the the explanatory variables and unknown regression coefficients multiplied by an arbitrary and unknown function of time.

The motivation behind the use of CPHM arises naturally, as one may be interested in which factors influence the survival. These factors, which can be one or various, may be categorical or continuous. For more complicated situations, the CPHM is a suitable regression-type model that incorporates the effect of each predictor on the shape of the survival curve [48].

Following Vidakovic [48], assuming that the log hazard for subject i can be modeled via a linear relationship:

$$\log h(t, x_i) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i}, \tag{2.6}$$

where $x_i = x_{1,i}, \cdots, x_{p,i}$ is the p-dimensional vector of covariates associated with subject i. When all covariates are equal to 0, Equation 2.6 corresponds to the log baseline hazard, i.e. $\log h_0(t) = \beta_0$. There also exists an alternative parametrization of the model where the baseline hazard can be set to correspond to a typical person for whom all covariates are averages of covariates from all subjects in the study. This is translated to the following modification of Equation 2.6:

$$\log h(t, x_i) = \log h_0(t) + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i}, \tag{2.7}$$

An exponentiation of Equation 2.7 leads to a more popular specification:

$$h(t, x_i) = h_0(t) \times \exp\{\beta_1 x_{1,i} + \cdots + \beta_p x_{p,i}\} = h_0(t) \times \exp\{x_i^T \beta\} \tag{2.8}$$

The name proportional refers to the fact that the ratio of hazards for two subjects, i and j, is free of t:

$$\frac{h(t, x_i)}{h(t, x_j)} = \exp\{(x_i^T - x_j^T)\beta\} \tag{2.9}$$

Lastly, it follows that the survival function $S(t)$ for a subject i can be written as:

$$S(t, x_i) = (S_0(t))^{x_i^T \beta} \tag{2.10}$$

where $S_0(t)$ is the survival function corresponding to the baseline hazard $h_0(t)$.

2.4 Application to Railway Wheelsets Maintenance

2.4.1 Problem Description and Assumptions

Data was acquired from a Portuguese railway company that transports passengers in a single line, and which extends 54 kilometers and serves 14 stations. A 17-year interval database, ranging from January 2001 up to July 2019, is considered. The company operates 18 electrical multiple units (EMUs). Each EMU has 4 cars and each car has eight wheels (i.e. four wheelsets). Figure 2.2 provides a schematic representation of a four-car unit.

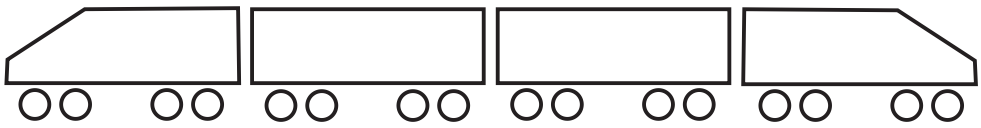

Figure 2.2: Schematic representation of a four-car unit

The present work assumes that inspections yield good estimates about the true state of the system and that maintenance actions have a deterministic outcome (i.e. they always restore the system to a pre-established level). MDP modelling considers:

(i). The wheel diameter (D), which is considered the main wheel profile indicator of the

lifecycle stage that a given wheel is at a certain epoch (n);

(ii). The wheel damage occurrence (such as rolling contact fatigue (RCF), flats or cavities), which is responsible for the most severe maintenance actions in railway wheelsets, shortening significantly their lifecycles;

(iii). The kilometers since last turning/renewal (kst) operation of each wheelset;

(iv). Three possible maintenance actions ($a = 1, 2, 3$):

- 'Do nothing' ($a = 1$): the wheelset is ok and it goes back to service in the same state;
- 'Renewal' ($a = 2$): the corrective or preventive maintenance actions would need to go beyond the scrap diameter, and so the wheel must be replaced by a new one;
- 'Turning' ($a = 3$): the wheelset goes to a turning lathe for its shape being replaced to values within the standards and it suffers a reduction/loss in its diameter.

The final objective of this modeling is to determine an optimal wheelset maintenance strategy based on wheel deterioration processes in an MDP framework. The maintenance costs, in the long run, are minimized and a decision map is provided depending on the wheel diameter, damage occurrence and kilometers since last turning/renewal. This decision map is expected to assign maintenance actions based on the condition of the wheelset, which may be used as a support for a train operating company defining its reprofiling policy.

This MDP is derived over an infinite planning horizon with a support of a software package [49] and the MDP is considered stationary, i.e. (i) the transition probabilities are assumed to be constant over time, and thus, the MTMs are independent of the epoch at which the transition occurs; (ii) the policy is independent of time.

The following 'State space' subsection explains the state space used in the modeling. Then, subsection 'Estimation of MTMs' discusses the estimation of the MTMs for each possible action. Subsection 'Reward/cost function' shows how the reward/cost functions are defined/estimated. The next subsection 'Optimal policy' provides the optimal maintenance policy, which maps the best possible action depending on the condition state of the wheelset. Finally, the last subsection discusses the main conclusions withdrawn from the analysis of the decision map obtained.

2.4.2 State Space

The state space will be defined based on three main chosen indicators for the wheelset states (i) wheel diameter (D), (ii) the kilometers since the last turning (kst) and (iii) the occurrence of damage. The wheel diameter varies from an initial diameter ($D_{initial}$) of 920 mm until a scrap diameter (D_{scrap}) of 850 mm and the diameter categories are discretized in intervals with amplitudes of 1 mm (i.e. 70 different levels). The kilometers since last turning/renewal (kst) vary from 0 up to 350 000 km in intervals of 10 000 km (i.e. 36 different levels/epochs). Finally, a wheelset can be in a state of damage or not, in a total of 70 states with damage, which are kept at the end of the state space. Transitions from damaged states to non-damaged states are compulsory because once the damage is detected, it must be removed, and hence, damaged states do not have the extension depending on the kilometres since last turning. Consequently, a total of 2590 different states, $s \in \{s_1, s_2, \cdots s_{2590}\}$, are defined.

2.4.3 Estimation of MTMs

An MTM has to be defined for each possible action. This section is divided into three subsections explaining the estimation of the 'Do nothing' MTM ($\boldsymbol{P}_1$), the 'Renewal' MTM ($\boldsymbol{P}_2$) and the 'Turning' MTM ($\boldsymbol{P}_3$).

As explained earlier, this study considers the wheel diameter (D) as the main indicator

of lifecycle stage of the wheel. In this analysis and as suggested in Figure 2.3, the wear in the wheelset – measured as the diameter loss due to wear (ΔD) – is assumed to be independent of the wheel initial diameter after a renewal or reprofiling. This assumption is reasonable as the hypothesis of independence cannot be rejected at a significance level of 0.05.

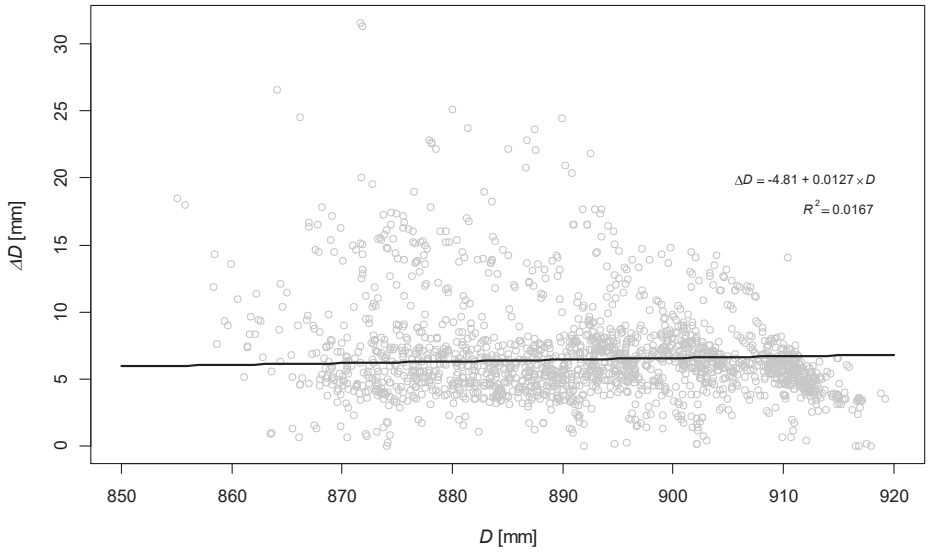

Figure 2.3: Diameter loss due to wear (ΔD) for different diameters (D)

Considering a homogenous Markov Chain, the transition matrix is decomposable into several sub-transition matrices – in a diagonal block form [50]. For a Markov Chain with a finite but large state space, the decomposition of the transition matrix tends to follow the most attractive approach. Having said that, the underlying problem of estimating MTMs can be divided into sub-problems that can be solved independently. This kind of approach will be followed in the next subsections with the estimation of sub-transition matrices.

'Do nothing' action ($a = 1$)

The 'Do nothing' action considers that the only possible way to increase the diameter of a wheel is through renewal. Furthermore, as data suggests, abrupt decreases in diameter (due to wear) are very unlikely to happen. Therefore, a simplification is considered where

the only possible transitions for a given state (not considering damage states transitions) is to move to a state of diameter immediately below (with probability θ) or stay in the same state (with probability $1 - \theta$), see Figure 2.4 for a schematic representation.

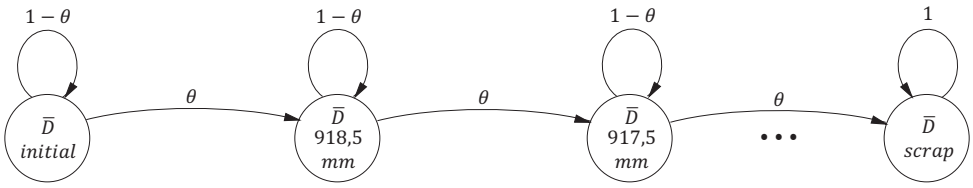

Figure 2.4: Transitions between states without damage depending on the parameter θ for the 'Do nothing' action

Regarding the diameter loss due to wear, it is possible to predict the mean value $\left(\overline{\Delta D}\right)$ with the Markovian approach each step of 10 000 km as:

$$\overline{\Delta D}_{(n)} = \boldsymbol{X}_0 \boldsymbol{P}^n \boldsymbol{\Delta D} \tag{2.11}$$

To derive these scalar mean values, the following variations in the wheel diameter for the wear states were considered:

$$\boldsymbol{\Delta D} = [0\ 1\ 2\ \cdots\ 69]^T\ mm \tag{2.12}$$

These were the possible variations for the wheelset diameters derived from the diameter state representative values considered in the sample, i.e. diameter categories mean values from $\bar{D}_{initial}$ =919.5 mm up to $\bar{D}_{scrap}$ =850.5 mm.

The initial state of the wheelset is:

$$\boldsymbol{X}_0 = [\Pr(\Delta D = 0)\ \Pr(\Delta D = 1)\ \cdots\ \Pr(\Delta D = 69)]\ mm \tag{2.13}$$

The reference value θ of the transition probabilities was obtained through a regression approach, considering a subset of the original data where no action (turning/renewal) was performed. Figure 2.5 shows the data points (in grey empty circles) which represent the

diameter loss (ΔD) due to wear for wheelsets without damage according to the variable "kilometers since last turning". A simple linear regression without an intercept was considered and, according to the ordinary least squares (OLS) criterion, the resulting regression line is represented in black (Figure 2.5). In fact, it is possible to show that considering the chosen transition matrix (with zeros in all entries that do not belong to the diagonal or upper diagonal), setting $\theta = 0.36$ for the values of ΔD (black cross in Figure 2.5) in Equation 2.12, solves Equation 2.11 considering the n-step transition probabilities (i.e. the probability that a process in state i will be in state j after n additional transitions). The assumption of not including an intercept (or in other words assuming that the intercept is equal to zero) is aligned with no wear, that is, $\Delta D = 0$, when a wheelset is new or just turned and it has no kilometres since last turning/renewal. Therefore, the value of $\theta = 0.36$ is chosen.

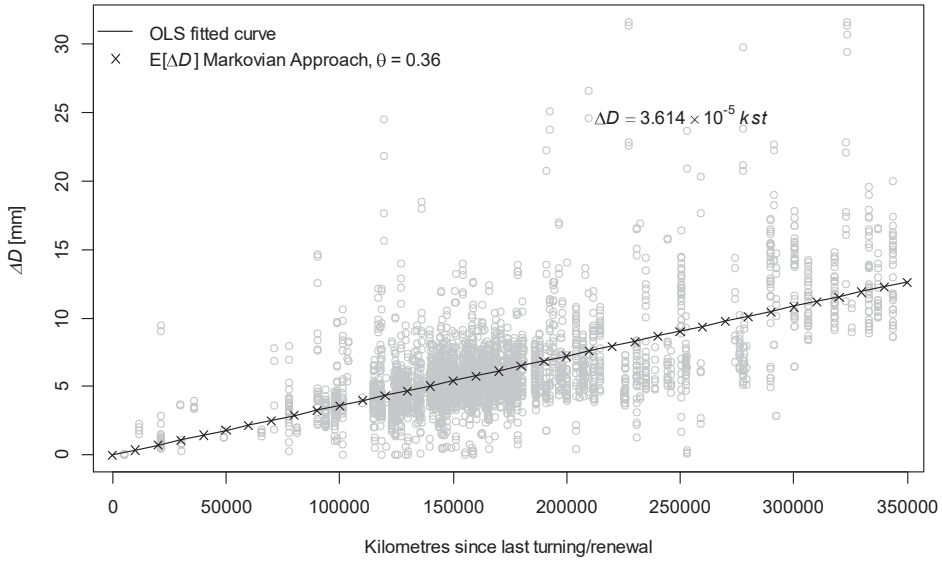

Figure 2.5: Diameter loss due to wear (ΔD) for wheelsets without damage with kilometers since last turning, applying Markovian approaches and linear regression without intercept

Next, transition probabilities to states with damage must also be derived. The main assumption is that a wheel stays damaged without a change in its diameter, since, in theory, once wheelset damage is detected the vehicle must be removed from service and the wheelset reprofiled. Therefore, transitions from wheels without damage to damaged states

are schematically represented in Figure 2.6.

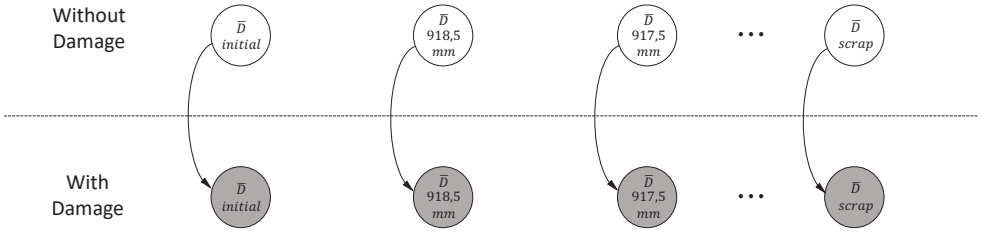

Figure 2.6: Considered transition probabilities to states with damage

For deriving the damage probabilities, a CPHM is implemented [47, 51]. Due to the database limitation of providing reliable information regarding the damage occurrence, another database from previous references [43, 52, 53, 54], and from a different train operating company is used for deriving the CPHM.

As discussed earlier, this model is a regression-type approach to survival curves, whenever the use of covariates (categorical or continuous) is needed. In this case study, this model is used for deriving the survival probabilities of a wheel given its diameter value. The hazard function $h(kst, D)$ in the CPHM for a wheelset (one observation) at a given value of $kst = k$ and $D = d$ can be calculated as:

$$h(k, d) = h_0(k) \times \exp\{\beta(d)\} \tag{2.14}$$

In the hazard equation above, the covariate is the tread diameter (D) and its coefficient β measures its size effect. The quantity $\exp\{\beta(d)\}$ is the hazard ratio linked to the covariate D, and it was shown to be statistically significant, with the upper term of the 0.95 confidence interval (CI) being 0.982 (slightly below 1), indicating that as the tread diameter (D) increases, the hazard decreases and, hence, length of survival increases, i.e. new wheelsets have longer survival than wheelsets whose diameter is close to the scrap diameter. Table 2.1 provides an overview of the CPHM output:

The cumulative hazard function $H(kst, D)$ in the CPHM for a wheelset (one

Table 2.1: CPHM output

Risk Factor	Estimate	p-value	Hazard Ratio (HR)	95% CI for HR
Diameter (D)	-0.022	$< 1 \times 10^{-16}$	0.978	$[0.975, 0.982]$

observation) at a given value of $kst = k$ and $D = d$ can be calculated as:

$$H(k, d) = H_0(k) \times \exp\{\beta(d)\} \tag{2.15}$$

In Equation 2.14 and Equation 2.15, $h_0(k)$ and $H_0(k)$ are baseline and cumulative baseline hazards, respectively, which are obtained when the value of the covariate d is set to 0 in the corresponding equations. Also, for a given wheel a general survival function is:

$$S(k, d) = (S_0(k))^{\exp\{\beta(d)\}} \tag{2.16}$$

Equation 2.16 shows how the Cox Model computes the survival probabilities based on a survival function $S_0(k)$ corresponding to the baseline hazard $h_0(k)$. The survival probability, for a fixed wheel diameter, at a given $kst = k$ represents the probability of survival beyond k, i.e.:

$$S(k) = \Pr(kst > k), \qquad k > 0 \tag{2.17}$$

Starting from the survival curves, it is more intuitive to understand the computation of the hazard rates. The hazard function at a fixed diameter, $h(kst)$, assesses the instantaneous risk of failure at $kst = k$, conditional on survival to that kst, or in mathematical notation:

$$h(k) = \lim_{\Delta k \to 0} \frac{\Pr(k < kst < k + \Delta k)}{\Delta k} = \frac{f(k)}{S(k)} = -\frac{S'(k)}{S(k)} \tag{2.18}$$

The equation above relates the hazard function to the survival function. For this work, hazard rates derived from the CPHM are displayed in Figure 2.7, where the line closer to the origin (with lowest hazard rates) corresponds to the hazard function for the highest

tread diameter representative value, $D_{initial}$ of 920 mm, and the upper curve corresponds to the hazard function for the lowest tread diameter representative value, D_{scrap} of 850 mm. The probability of occurring damage in a wheel of a given diameter at a certain kst is taken as simply the discretized values of hazard curves in Figure 2.7.

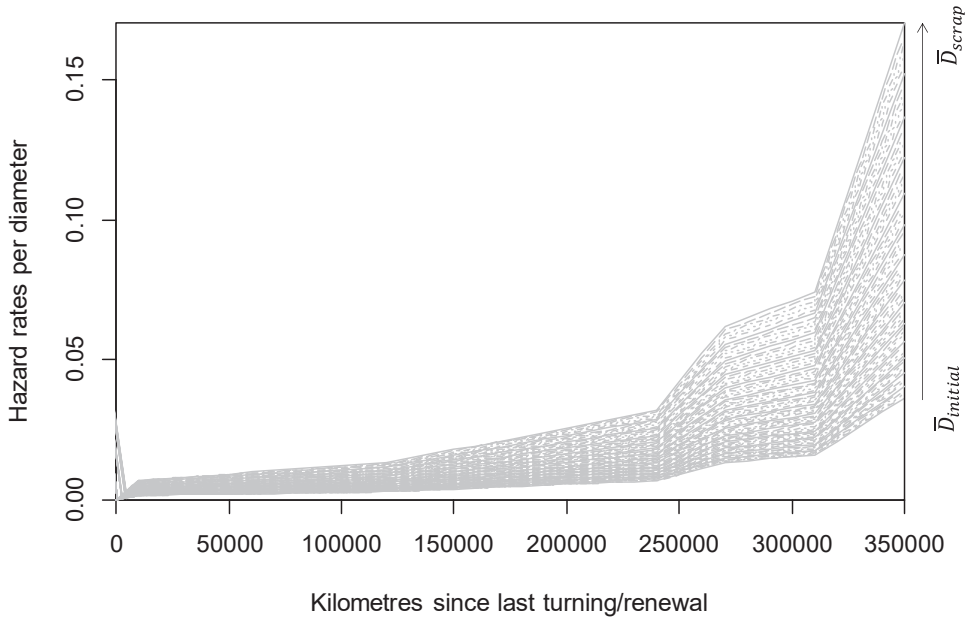

Figure 2.7: Estimated survival probabilities per diameter group representative values with kilometers since last turning

A sub-transition matrix ($\boldsymbol{P}_D$) for the damage probabilities, considering all 2520 states

without damage to the 70 states with damage can be represented in the following way:

$$\boldsymbol{P}_D \equiv [p^1_{i,j}] = \left[\begin{array}{cccc|l|l}
p(damage) & 0 & \cdots & 0 & \bar{D}_{initial} & \\
0 & p(damage) & 0 & \vdots & & \\
\vdots & 0 & p(damage) & 0 & & \textit{(kst 0 km)} \\
0 & \cdots & 0 & p(damage) & \bar{D}_{scrap} & \\
\hline
p(damage) & 0 & \cdots & 0 & \bar{D}_{initial} & \\
0 & p(damage) & 0 & \vdots & & \\
\vdots & 0 & p(damage) & 0 & & \textit{(kst 10k km)} \\
0 & \cdots & 0 & p(damage) & \bar{D}_{scrap} & \\
\hline
p(damage) & 0 & \cdots & 0 & \bar{D}_{initial} & \\
0 & p(damage) & 0 & \vdots & & \\
\vdots & 0 & p(damage) & 0 & & \vdots \\
0 & \cdots & 0 & p(damage) & \bar{D}_{scrap} & \\
\hline
p(damage) & 0 & \cdots & 0 & \bar{D}_{initial} & \\
0 & p(damage) & 0 & \vdots & & \textit{(kst 350k km)} \\
\vdots & 0 & p(damage) & 0 & & \\
0 & \cdots & 0 & p(damage) & \bar{D}_{scrap} &
\end{array}\right] \tag{2.19}$$

Transition to damage is considered independent of wear transitions (ΔD); hence, it follows that the joint transition probability of damage and wear is equal to the product of the marginal transition probabilities, as follows:

$$\Pr(wear \cap damage) = \Pr(wear) \cdot \Pr(damage) \tag{2.20}$$

Therefore, the subtransition matrices for wear need to be modified so that probabilities of damage are incorporated. The non-zero entries of these matrices, represented as $\boldsymbol{P}_W$, are

computed as:

$$\begin{cases} p_{i,i+70} = (1-\theta)(1-p(damage)); & i = j + 70k \\ p_{i,i+71} = \theta(1-p(damage)); & i = j + 70k, \quad j = 1, 2, \cdots, 69, \quad k = 0, 1, \cdots, 34 \\ p_{i,i+70} = 1 - p(damage); & i = 70(k+1) \\ p_{i,i} = 1 - p(damage); & i = 2521, 2522, \cdots, 2590 \end{cases} \tag{2.21}$$

In a matrix form, the final MTM for the 'Do nothing' action ($\boldsymbol{P}_1$) is a 2590 by 2590 matrix composed by the sub-transition matrices in a diagonal form, as follows:

$$\boldsymbol{P}_1 \equiv \left[p_{i,j}^1\right] = \left[\begin{array}{ccccccc:c} \text{kst} & \text{kst} & \text{kst} & & \text{kst} & \text{kst} & \text{states} & \\ \text{0 km} & \text{10 km} & \text{20 km} & \cdots & \text{340 km} & \text{350 km} & \text{with damage} & \\ \hline \mathbf{0} & \boldsymbol{P}_W^{70\times70} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & \text{0 km} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{P}_W^{70\times70} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & \text{10 km} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \boldsymbol{P}_D^{2520\times70} & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{P}_W^{70\times70} & \mathbf{0} & & \text{340 km} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{P}_W^{70\times70} & & \text{350 km} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I}^{70\times70} & \text{w/ damage} \end{array}\right]^{2590\times2590} \tag{2.22}$$

'Renewal' action ($a = 2$)

Concerning the 'Renewal' action, regardless of the current state of the wheel (damaged or undamaged), transitions to the initial state are assumed to be certain, as described in Figure 2.8.

Therefore, the MTM for the 'Renewal' situation ($\boldsymbol{P}_2$) is a 2590 by 2590 matrix as

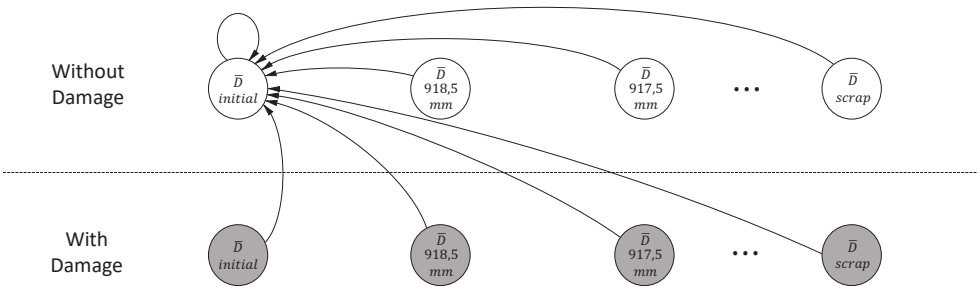

Figure 2.8: Transitions between states for the 'Renewal' action

follows:

$$\boldsymbol{P}_2 \equiv \left[p_{i,j}^2\right] = \begin{bmatrix} \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}^{2590\times 2590} \tag{2.23}$$

'Turning' action ($a = 3$)

The possible transitions between states for the 'Turning' action are schematically represented in Figure 2.9.

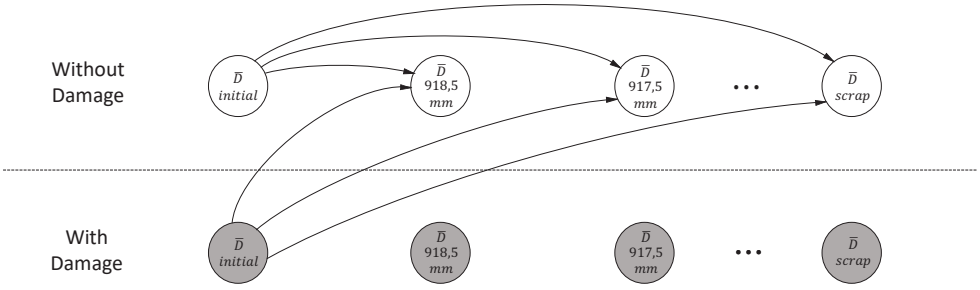

Figure 2.9: Transitions between states for the 'Turning' action

When turning a wheelset, there is a distinct loss in the diameter due to turning (reprofiling of the wheel) if it is a situation of correcting damage or if it is a situation of preventive turning. In the case of correcting damage, the diameter loss tends to be significantly larger on average and with a higher dispersion as depicted in Figure 2.10.

The Portuguese train operating company database does not distinguish between turning situations with damage from without damage. Due to this database limitation, the diameter

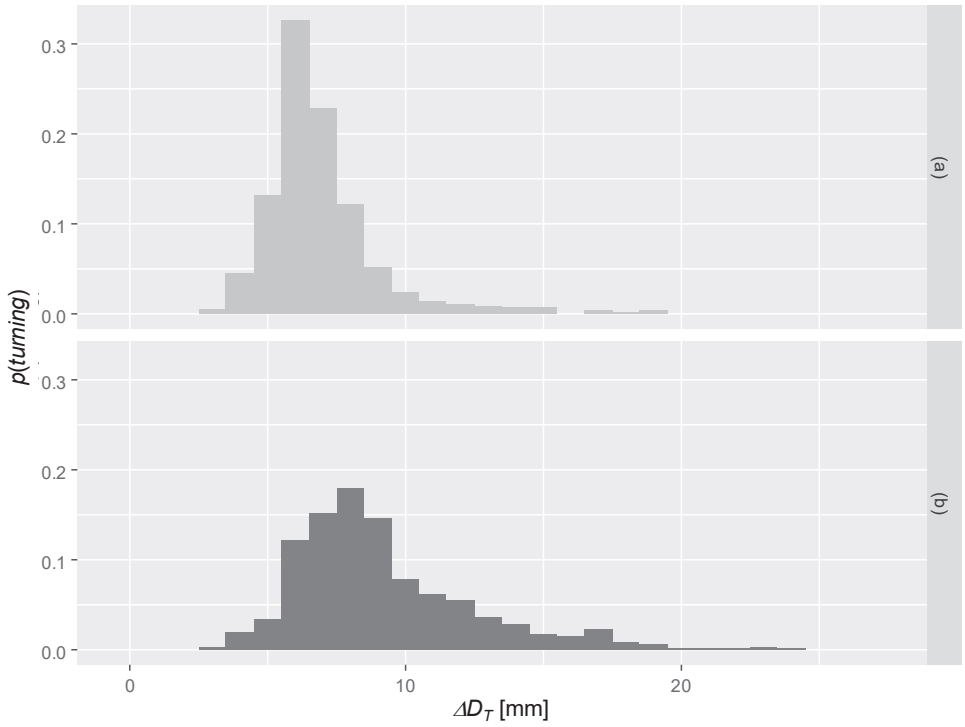

Figure 2.10: Histograms of the loss in diameter due to turning (ΔD_T) in a wheelset: (a) without damage and (b) with damage

loss due to turning (ΔD_T) probability distributions for damaged and undamaged wheels was considered the same as in Braga and Andrade [43].

The probabilities represented in Figure 2.10 withdrawn from Braga and Andrade [43] were calculated using the relative frequency from past samples as an approximation of the transition probabilities, that is,

$$p(turning) = \frac{n_j}{N} \tag{2.24}$$

In which n_j is the number of wheelsets that transit to a class j of diameter loss and N is the total number of wheelsets. The MTM for the 'Turning' action assumes that the transitions to next states are limited, meaning that a transition from a state to another one with a great loss in the diameter does not happen at some point (according to Figure 2.10, 30 mm is defined as the maximum loss in diameter possible).

Therefore, regarding transitions from one state to another, the probabilities are

composed by zeros to states before the current one and zeros for states after the current one that the 'Turning' action can not reach.

When a wheelset is turned, it goes back to a state where kilometers since last turning/renewal (kst) are zero, and if it has damage it goes to a state without damage, since once the damage is detected it must be removed. As it is not possible to turn a wheelset beyond the scrap diameter, when the wheelset is in a scrap diameter state, at some point of its kilometers since last turning (kst), and the histograms of Figure 2.10 (a) indicate diameter losses that go beyond the scrap diameter for that final state, the probabilities of the remaining transitions are summed up becoming the probability value for the wheelset to stay at the final state, i.e. the scrap diameter. Having said that, it is possible to compose the sub-transition matrix for the 'Turning' action from states without damage ($\boldsymbol{P}_{TND}$):

$$\boldsymbol{P}_{TND} \equiv \left[p^3_{i,j}\right] = \begin{bmatrix}
p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & p_{1,3}^{3^{TND}} & \cdots & p_{1,30}^{3^{TND}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & p_{1,2}^{3^{TND}} & \cdots & p_{1,29}^{3^{TND}} + p_{1,30}^{3^{TND}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p_{1,1}^{3^{TND}} & \cdots & p_{1,28}^{3^{TND}} + p_{1,29}^{3^{TND}} + p_{1,30}^{3^{TND}} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \sum_{j=1}^{30} p_{1,j}^{3^{TND}}
\end{bmatrix} \tag{2.25}$$

In the same way, using now the probability values withdrawn from Figure 2.10 (b), it is possible to compose the sub-transition matrix for the 'Turning' action from states with

damage ($\boldsymbol{P}_{TD}$):

$$\boldsymbol{P}_{TD} \equiv \left[p^3_{i,j}\right] = \begin{bmatrix}
p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & 0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & 0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} & 0 & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & 0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} & 0 & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} & 0 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & p^{3^{TD}}_{1,3} & \cdots & p^{3^{TD}}_{1,30} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p^{3^{TD}}_{1,1} & p^{3^{TD}}_{1,2} & \cdots & p^{3^{TD}}_{1,29} + p^{3^{TD}}_{1,30} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & p^{3^{TD}}_{1,1} & \cdots & p^{3^{TD}}_{1,28} + p^{3^{TD}}_{1,29} + p^{3^{TD}}_{1,30} \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \sum_{j=1}^{30} p^{3^{TD}}_{1,j}
\end{bmatrix} \tag{2.26}$$

Finally, the MTM when the 'Turning' action is chosen ($\boldsymbol{P}_3$) is composed as follows:

$$\boldsymbol{P}_3 \equiv \left[p^3_{i,j}\right] = \left[\begin{array}{cccccc:c}
\text{kst} & \text{kst} & \text{kst} & & \text{kst} & \text{kst} & \text{states} & \\
\text{0 km} & \text{10 km} & \text{20 km} & \cdots & \text{340 km} & \text{350 km} & \text{with damage} & \\
\hdashline
\boldsymbol{P}^{70\times70}_{TND} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \text{0 km} \\
\boldsymbol{P}^{70\times70}_{TND} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \text{10 km} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\boldsymbol{P}^{70\times70}_{TND} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \text{340 km} \\
\boldsymbol{P}^{70\times70}_{TND} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \text{350 km} \\
\boldsymbol{P}^{70\times70}_{TND} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \text{w/ damage}
\end{array}\right]^{2590\times2590} \tag{2.27}$$

Figure A.1, Figure A.2 and Figure A.3 attached in the Appendix provide useful diagrams that detail the state space and the possible transitions from one state to another for each of the three possible actions, $a = 1, 2, 3$, respectively.

2.4.4 Cost Function

As the MDP Toolbox (MATLAB ® software) chosen to solve this problem used a reward maximization function to derive the expected total discounted value rewards, the values used to represent the costs of the maintenance operations must be negative [49]. To derive the reward/cost function, a reward vector (q) for each action chosen ($a = 1, 2, 3$) is specified.

It is assumed that the 'Do nothing' action ($a = 1$) does not hold any operational cost. However, it is important to guarantee, due to the state space adopted constraints, that when the wheelset reaches states with a diameter equal to the scrap diameter, kilometers since last turning/renewal (kst) of 350 000 km or damaged states, other option different from 'Do nothing' is not chosen. This is done by setting at these critical states cost values larger than the ones used in the remaining actions. For these states, it was assumed that the values of

10000 € should be assigned as follows:

$$q_i^1 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ q_{70}^1(s_{70}) \\ 0 \\ \vdots \\ 0 \\ q_{140}^1(s_{140}) \\ 0 \\ \vdots \\ 0 \\ q_{140+70}^1(s_{140+70}) \\ 0 \\ \vdots \\ 0 \\ q_{2450}^1(s_{2450}) \\ q_{2451}^1(s_{2451}) \\ \vdots \\ q_{2520}^1(s_{2520}) \\ q_{2521}^1(s_{2521}) \\ \vdots \\ q_{2590}^1(s_{2590}) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ -10000 \\ 0 \\ \vdots \\ 0 \\ -10000 \\ 0 \\ \vdots \\ 0 \\ -10000 \\ 0 \\ \vdots \\ 0 \\ -10000 \\ -10000 \\ \vdots \\ -10000 \\ -10000 \\ \vdots \\ -10000 \end{bmatrix} \begin{array}{l} \\ \\ \\ \rightarrow \textit{Scrap diameter} \\ \\ \\ \\ \rightarrow \textit{Scrap diameter} \\ \\ \\ \\ \rightarrow \textit{Scrap diameter} \\ \\ \\ \\ \rightarrow \textit{Scrap diameter} \\ \rightarrow \textit{kst = 350k km} \\ \vdots \\ \rightarrow \textit{kst = 350k km} \\ \rightarrow \textit{damage} \\ \vdots \\ \rightarrow \textit{damage} \end{array} \tag{2.28}$$

For the 'Renewal' action ($a = 2$), a value of -800 € is set, regardless of the state a wheelset is; hence, the reward vector is as follows:

$$q_i^1 = \begin{bmatrix} q_1^2(s_1) \\ \vdots \\ q_{2590}^2(s_{2590}) \end{bmatrix} = \begin{bmatrix} -800 \\ \vdots \\ -800 \end{bmatrix} \tag{2.29}$$

Turning a wheelset without damage (wheel states without damage) is set as having a cost of 50 € while doing turning for correcting a damaged wheelset (wheel states with damage) is set as having a cost of 150 € . However, there are some critical states where a

'Renewal' action is needed. Those cases are the ones when the scrap diameter is reached and, for the MDP does not "choose" a 'Turning' action but instead a 'Renewal' action, they are "penalized" with a cost of 10000 €. Summing up, the reward vector for the 'Turning' action is as follows:

$$q_i^3 = \begin{bmatrix} q_1^3(s_1) \\ \vdots \\ q_{69}^3(s_{69}) \\ q_{70}^3(s_{70}) \\ q_{71}^3(s_{71}) \\ \vdots \\ q_{139}^3(s_{139}) \\ q_{140}^3(s_{140}) \\ q_{141}^3(s_{141}) \\ \vdots \\ q_{140+69}^3(s_{140+69}) \\ q_{140+70}^3(s_{140+70}) \\ q_{140+71}^3(s_{140+71}) \\ \vdots \\ q_{2449}^3(s_{2449}) \\ q_{2450}^3(s_{2450}) \\ q_{2451}^3(s_{2451}) \\ \vdots \\ q_{2520}^1(s_{2520}) \\ q_{2521}^1(s_{2521}) \\ \vdots \\ q_{2590}^1(s_{2590}) \end{bmatrix} = \begin{bmatrix} -50 & \\ \vdots & \\ -50 & \\ -10000 & \rightarrow \textit{Scrap diameter} \\ -50 & \\ \vdots & \\ -50 & \\ -10000 & \rightarrow \textit{Scrap diameter} \\ -50 & \\ \vdots & \\ -50 & \\ -10000 & \rightarrow \textit{Scrap diameter} \\ -50 & \\ \vdots & \\ -50 & \\ -10000 & \rightarrow \textit{Scrap diameter} \\ -10000 & \rightarrow \textit{kst = 350k km} \\ \vdots & \vdots \\ -10000 & \rightarrow \textit{kst = 350k km} \\ -150 & \rightarrow \textit{damage} \\ \vdots & \vdots \\ -150 & \rightarrow \textit{damage} \end{bmatrix} \tag{2.30}$$

It is important to mention that ideally such a cost function would include the operational costs/impacts of using the wheelset in such condition, namely noise impact, wear and damage impact in the rail component, passenger comfort, safety impacts, reliability impacts, etc. Nevertheless, these more comprehensive impacts are hard to estimate and thus are not included in the present cost function.

2.4.5 Optimal Policy

An optimal policy is computed using the MDP Toolbox (MATLAB® software). The optimal policy of the decision process associated with maintaining a railway wheelset can then be organized in a graphic table for all states (damaged and undamaged) with the evolution of the kilometers since last turning/renewal (kst), as shown in Figure 2.11.

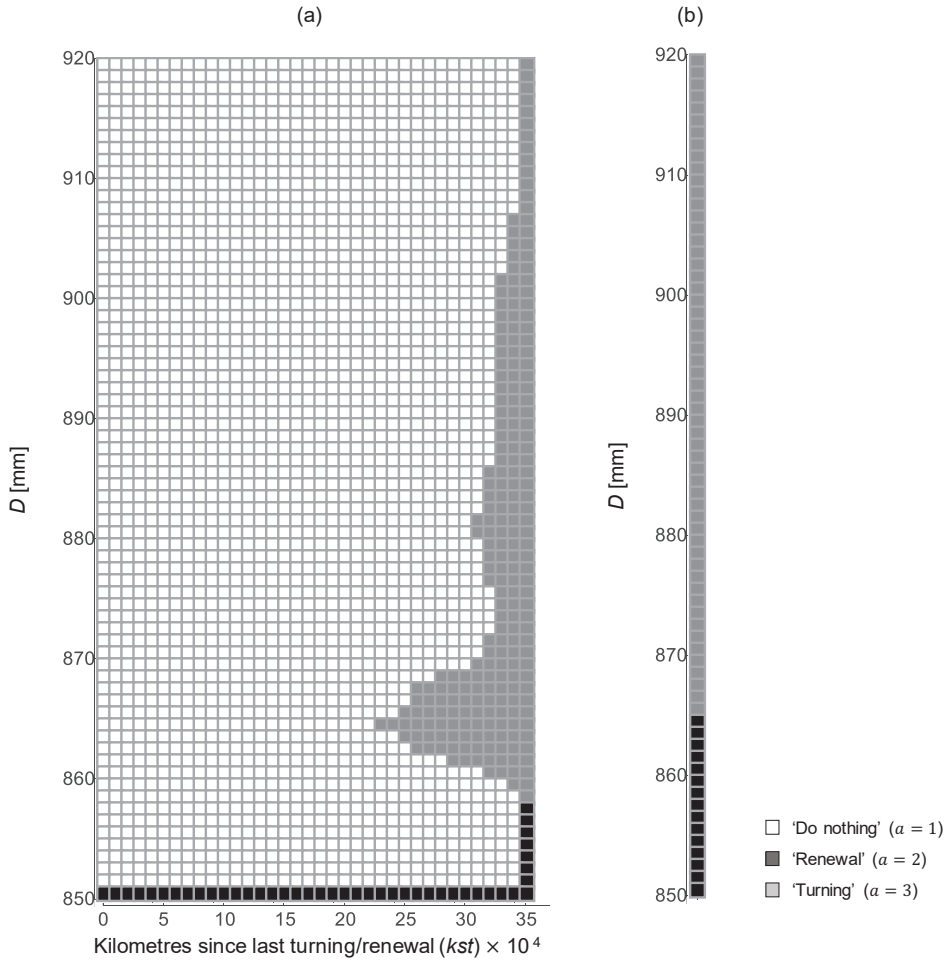

Figure 2.11: Map of decisions for wheelsets with (b) and without (a) damage with the evolution of the kilometres since last turning (kst)

By analysing Figure 2.11, one can see that the transition probability methods adopted and reward values were chosen in the section 'Reward/cost function' resulted in actions that were intended (a) for the undamaged wheelsets and (b) for damaged wheelsets. For the last, Figure 2.11 (b) shows that only actions of 'Turning' or 'Renewal' are assigned,

being the 'Renewal' actions recommended for the last states where the 'Turning' action would go beyond the scrap diameter.

For the undamaged wheelsets, Figure 2.11 (a) indicates that the recommended actions change with kilometers since last turning (kst) and wheelset diameter (D). It helps to define a critical point, say k^*, as being the lowest kst such that turning action is recommended (here on coordinates kst = 240 000 km and D = 864 mm. The grey pattern, which covers part of the right side, suggests that turning actions should be performed earlier as the diameter decreases. This relationship holds until approximately the critical point k^*. For diameters below k^* (864 mm or closer to scrap diameter), the strategy shifts to allowing more kst. Indeed, for diameters below 858 mm, the best strategy is to not perform turning at all and let the wheelset wear out until scrap diameter. Hence, if one imagines a line connecting k^* to the first grey square on top (kst = 350 000 km and D = 920 mm) and a second line connecting k^* to the last grey square on bottom (kst = 350 000 km and D = 858 mm), it is evident that both signs and absolute values of the corresponding slopes are different. For newer wheelsets until about the point D = 870 mm, there is a slow decrease of kst while the decrease in D occurs at a higher rate, meaning that those wheels can support long periods without undergoing 'Turning' actions. Then, from D = 870 mm down to k^* diameter, there is a fast decrease of both kst and D. Below k^* diameter, i.e. wheelsets whose diameters are close to scrap diameter, the "slope" changes sign and there is a fast increase of kst as D decreases.

Therefore, Figure 2.11 serves as a guideline for condition-based maintenance, that is, depending on the diameter (D), kilometers since last turning (kst) and whether or not damage has occurred, it provides the optimal action that minimizes the total costs for each defined wheelset state. However, such policy may not be effective in practice, as this requires train operating companies to have exceptional maintenance management and control over their assets, which might be unrealistic. Therefore, a modification of the policy, making it vary across only one parameter, in this case, kilometers since last turning

(kst), can be compared with the expected cost of the optimal policy so that an "easy-to-apply" maintenance strategy that has cost closest to the optimal can be implemented.

For this new strategy, some entries of the undamaged policy in Figure 2.11 (a) are modified according to the following rule: for a fixed "kilometers since last turning" (kst) value, all squares at that column or before that value's column will be filled with 1's (corresponding to action 'Do nothing'), all the squares after that value's column will be filled with 3's (corresponding to action 'Turning'), with the exception being the squares in black (action 'Renewal'), which will remain the same as in the optimal policy. Therefore, with the exception of the renewal squares, all the squares across the same value of "kilometers since last turning" (kst), i.e. all the entries on the same column, will have either 1 or 3, independently of the diameter (D). For example, if the new strategy sets the turning action to be performed after 150 000 km kst, then all squares before and at column 150 000 km kst will be filled with 1, or color white, and all other squares will be filled with 3, or color grey, (except the ones originally marked with 2, which remain black).

Under this framework, 21 different values of kst were used to build new policies and had their expected long-run cost extracted (no changes were made to the transition probability matrices or cost vectors, inputs of MDP approach). The cost results for these 21 different policies can be compared with the long-run expected cost arising from the optimal policy displayed in Figure 2.11. To facilitate the comparisons, the optimal solution cost is set to 1, and all other policies' costs (which are higher) are displayed as percentages of increase compared to the optimal one, as shown inFigure 2.12.

Figure 2.12 compares the results of the different kst cut-off policies for 21 different values (empty circles) of kilometers since last turning, namely from 150 000 km to 350 000 km in steps of 10 000 km, to the long-run average cost of the optimal policy (displayed as the solid line). The optimal policy is a line, not a point, since there is no kst cut-off as it considers a policy involving both kst and D. The optimal policy has the lowest expected long-run cost, followed by the policy which sets the cut-off for kst as

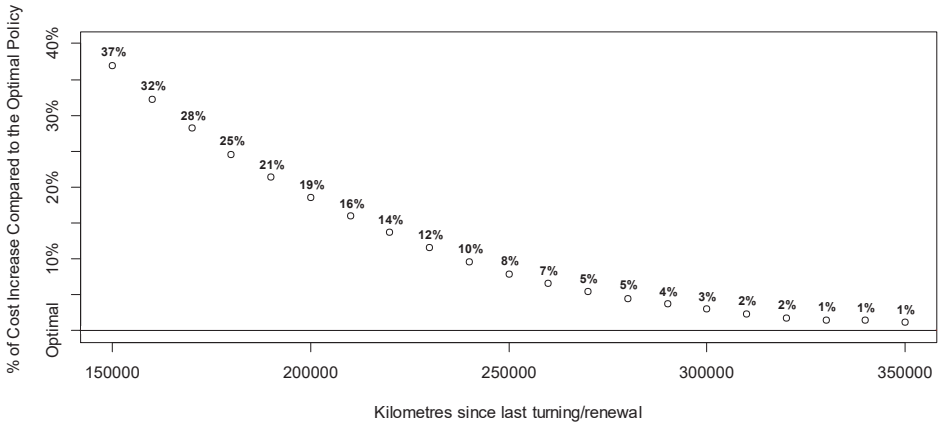

Figure 2.12: Comparison of optimal policy cost with costs arising from different policies based on kst cut-offs

350 000 km, which has an associated average cost about 1.1% higher than the optimal one. Policies with cut-offs in the neighborhood of 350,000 km, i.e., 330 000 km and 340 000 km, have also similar costs, about 1.5% and 1.4%, respectively, higher than the optimal. In general, policies with cut-offs inside the interval $]300\,000$ km,350 000 km$]$ have associated long-run average cost within 2.5% of the optimal policy's cost and, hence, perform fairly good.

2.5 Conclusions and Further Research

A data-driven model based on the MDP approach was implemented in order to provide the train operating company with a better decision-making process in terms of the turning of wheelsets and replacement policy. The change in tread diameter, kilometers since last turning/renewal and damage occurrence were used to define a discrete state space with a total of 2590 states. A set of 3 possible actions were considered to account for all possible decisions that can be undertaken after a wheelset is measured, namely: (1) 'Do nothing', (2) 'Renewal' and (3) 'Turning'. Reasonable values for the cost vectors were set and optimization in terms of minimization of costs was performed with the support of the MDP Toolbox [49].

The main result of this work comes in the form of a map of decisions, in terms of

actions to be taken for each discretization of kilometers since last turning/renewal (kst) and wheelset diameter (D), where two distinguished wheelset policy paths are easily identified. These two paths are separated by a critical point k^*, as tipping point for turning decisions: up to this point (situated at 230 000 km kst and with wheel diameter of around 865 mm) predictive turning actions have increasing importance in terms of extending the wheelset lifecycle, whereas, after this critical point, predictive turning starts progressively losing importance in the wheelset reprofiling policy.

Going deeper and focusing on the turning silhouette of the decision map, actually three main trends are suggested: for wheelsets in earlier stages of lifecycle (until about 870 mm of diameter), there is a slow decrease of kst as D decreases at a higher rate, hence, turning actions can be postponed to higher kst without the need of constant measurements of the diameter (the slack values for diameter lie within a big interval, in this case ranging from 920 mm down to 870 mm, or a 50 mm slack). Then, from D = 870 mm down to k^* diameter of 864 mm, there is a fast decrease of both kst and D, so efficient and more regular monitoring on the values of the diameter would be necessary to guarantee that turning actions would be performed at the optimal configuration. Finally, for wheelsets whose diameters are less than the k^* diameter (of 864 mm), or close to the scrap diameter, there is a fast increase of kst as D decreases, and the recommendation is to allow the wheelset to wear out until scrap diameter, as many frequent inspections would be necessary to guarantee that turning action would still be captured in an "optimal" setting, and hence, it is unlikely that turning at this point would be cost-effective.

Another contribution of this work concerns the ease of maintenance policy implementation. As a policy contemplating two parameters might be hard to implement in the context of a train operating company, it is interesting to compare the optimal results with those obtained from simpler policies. Hence, policies based on 21 different cut-offs values for kilometers since last turning were tested so that, for a given value of the cut-off (e.g. 250 000 km), all actions before or on that mark would correspond to 'Do nothing' and

all actions after that mark would result in 'Turning'. The exception would be the actions assigned 'Renewal' in the optimal policy, which were kept the same. In this framework, no consideration of the diameter would have to be made by the maintenance team in order to decide between 'Turning' or 'Do nothing', as kst would be the only input needed. Results displayed in Figure 2.12 revealed that, although the optimal solution had the lowest expected long-run cost, a good strategy would be to set the cut-off for kst as close as possible to 350 000 km, independently of the diameter. This strategy would increase the average long-run cost by only about 1.1%. In general, policies with cut-offs within the interval]300 000 km,350 000 km] perform fairly good and are recommended.

By inspecting the different cost values in Figure 2.12, it is not clear that, in the presence of states for which kst would exceed 350 000 km, the optimal configuration would lead to higher values of kst associated with turning recommendation. In fact, considering the current policy of the train operating company, even the recommendation of turning at 350 000 km translates into doubling the amount of kst that is currently established in the maintenance strategy. In this context, investigating further the kst of 350 000 km may be unrealistic, hence the limited state space based on maximum kst of 350 000 km is taken as one of the limitations of this study. As a second limitation, although the state space described a range of different configurations involving kilometers since last turning/renewal, wheel diameter and occurrence of damage, the evolution of other important variables such as the flange thickness and height as well as the angle inclination is not controlled. The goal of this study was to look at the problem of maintenance planning from the perspective of immediate action cost-optimization, and this is reflected in the MDP model objective function. As a secondary goal, this research work aimed to provide a reprofiling policy for wheels that was also "easy to implement". To accomplish this, the analysis was limited to the most influential parameters in terms of life cycle cost minimization, although inclusion of other parameters may be important if the primary goal is to look at the problem from a safety perspective, which may also limit the optimal

decision map suggested.

Therefore, the inclusion of those parameters is suggested as future research. The decision-making for this type of maintenance approach is mostly driven by wheel profile measurements. In this chapter, wheel profile observations are considered to be ”perfect”. However, uncertainties in the measurements should also be taken into account and they have different behaviors depending on the measurement procedure/device. This motivate the subsequent chapter 3, which is based in the paper [19]. Therefore, incorporating these different uncertainties in the modeling of the MDP and assessing their impact in the optimal values are also recommended topics for further research. Finally, this type of modeling to support maintenance decisions can also be integrated with other maintenance scheduling models [55], for improving service availability, reliability and logistic costs within train operating companies.

CHAPTER 3

UNCERTAINTY AROUND RAILWAY WHEELSET INSPECTIONS

3.1 Introduction

Data acquisition plays a decisive role in determining cost-effective maintenance policies for railway wheelsets. Wheelsets deteriorate with usage, and a good maintenance strategy should keep their condition at an acceptable level by monitoring various important parameters and ensuring they are under control. At the same time, accurately measuring those parameters is crucial, so that turning and replacement actions can be undertaken very closely to cost-effective targets, resulting in minimal costs of losing useful life of wheels.

In terms of monitoring of railway wheels, Alemi et al. [4] point out the differences in terms of data acquisition. In-service and in-workshop inspection, wayside and onboard measurement and diagnostic and prognostic approaches are some examples. In this chapter, a comparison between data acquired from manual (gauge device), laser device and under-floor wheel lathe measurements is presented for three wheelset parameters: flange thickness (Ft), flange height (Fh) and flange slope (qR). Figure 3.1 provides a schematic representation of the three variables which are measured relative to three fixed measurements (a, b, c) and a tread datum position (T) located 70 mm from the flange back.

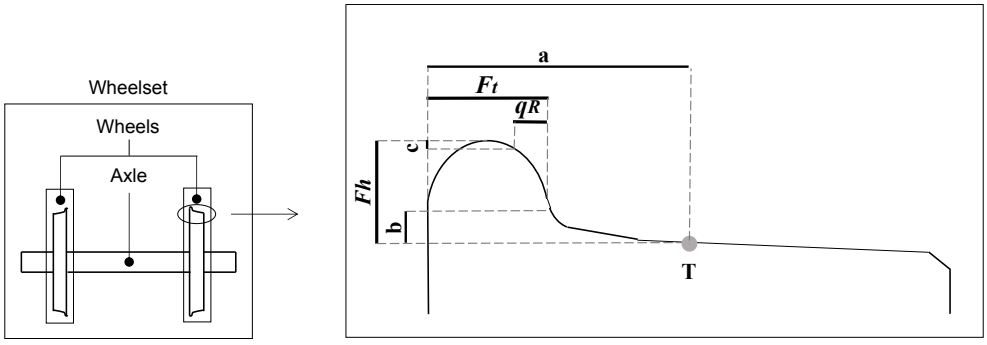

Figure 3.1: Railway wheelset on the left and measurements of its wheel profile on the right: flange thickness (Ft), flange height (Fh) and flange slope (qR)

The main interest of this study lies in the comparison of precision of these measurements. Intuitively, the distribution of each type of measurement (for each parameter) should be centred around some value, and this value should be the same across the different types of measurements. Therefore, the variance around each same value should be analysed to infer whether it is different for manual, laser and/or under-floor wheel lathe measurements. In other words, the research question is: ”do the three methods of measurement agree statistically?”. An additional challenge is that these three technologies were not used complementary, i.e., the wheelsets were not measured by the three methods at the same time, making direct pairwise comparisons not possible, which poses additional challenges to answer the former question.

To motivate the research question, this paper bases its analysis on a real situation of the Portuguese train operating company from chapter 2. This company needs to monitor the condition of the profiles of their train wheels periodically, since the measurements of Fh, Ft and qR need to be within safety limits and they change along time with the vehicle utilization and due to the contact of the wheels with the rails.

These periodic inspections have traditionally been executed through the use of a gauge device (manual inspection), though since 2017 it has slowly been replaced by a laser device. Measurements taken with the laser device tend to be faster than with gauge devices, taking approximately 20 minutes for a multiple unit with 16 wheelsets, which compares with an average of 90 minutes for manual inspection with a gauge device. These time savings are easily converted into cost savings and have been the economic justification to make the investment of buying the laser equipment (a more expensive inspection device). Moreover, gauge devices are theoretically more prone to human errors, and they tend to have a lower precision than the laser device.

Parallelly, if the wheels are found to be out of the safety limits in the inspections, they will have to be reprofiled in a turning machine to safer values. In this operation, the train goes into an under-floor turning wheel lathe machine, which also measures the

wheel profile parameters through a Computer Numerical Control (CNC) machine. After a reprofiling operation, there is a diameter loss in the wheel. The measurements of the opposite side wheel, the measurements of the wheels of the same bogie and wheel between bogies have also to be checked, since they must also not vary much between each other. Any other wheel of the vehicle that is out of these specifications relative to the turned one also have to be reprofiled. Since the train is already out of service for maintenance and inside the under-floor wheel lathe, technicians usually check the other wheels and make the decisions of turning them or not, using the CNC machine (and not the gauge, nor the laser device). Therefore, measuring wheels through this procedure is crucial for the decision-making of turning or not a wheel and directly affects its life-cycle. Note that wheels may also be reprofiled for removing a damage on their surface, a situation where the specifications after turning a wheel still stand.

Figure 3.2 depicts the three types of procedures for measuring wheels. Figure 3.2 (a) shows a gauge device, where the measurements are withdrawn and read in a similar way of a Vernier calliper with the device tips touching the wheel surface. At the top of Figure 3.2 (b), there is a technician performing a turning operation (who is checking if a damage is completely removed from the wheel surface) and below this image one can see the measurements of the wheel profile displayed on the screen of the CNC machine. Finally, Figure 3.2 (c) shows the mechanisms of measuring a wheel profile using the laser device. This is a contact-free technology that evaluates all the wheel profile lengths. The operator approaches to the wheel surface a scan device that emits three laser lines responsible for processing an image of the wheel shape on a portable screen. Once the wheel body contour is totally defined on the screen, the information system indicates if the wheel values are within or beyond the tolerance limits.

At the bottom of Figure 3.2, there are visual examples of the distributions of these measurement procedures. The assumption is that the underlying distributions are approximately normal and there is no disagreement due to bias, so that the three distribution

plots are centered around a common mean value (equal to zero). Then, the differences observed in each of the bell curve shapes are due to the differences in the variances. Knowing that the laser device has the best precision ($\pm$ 15 µm) followed by the CNC of the wheel lathe and the gauge device both having a lower precision ($\pm$ 0.05 mm), an intuitive expected behaviour of the distributions should look like the ones displayed in Figure 3.2.

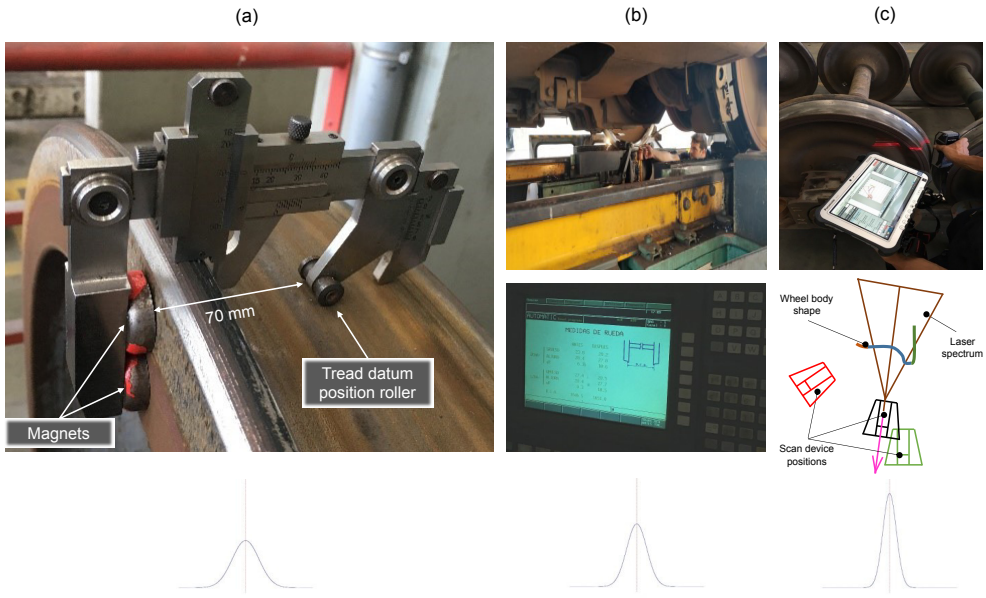

Figure 3.2: Different railway wheel profile measurement procedures and examples of hypothetical distributions at the bottom: (a) manual gauge device; (b) CNC under-floor turning lathe machine; (c) laser device

If data exhibited the behavior at the bottom of Figure 3.2, it would be possible to quantify, in terms of financial savings, the differences in adopting one technology in lieu of the other(s). For example, Madanat and Ben-Akiva [56] provide a comparison between two models based on Markov decision processes: in the first model random errors are explicitly considered in the inspection measurements, and the second model uses the commonly adopted assumption that inspections are perfect, i.e., they reveal the true condition state of the system being monitored. The main result shows that the minimum expected life cycle cost increases with an increase in the measurement uncertainty or, in other words, the higher the standard deviation associated with the measurement error, the higher the life cycle costs.

3.1.1 Measurement Error

Measurements may be influenced by a number of factors. They may be influenced by the way each operator uses the measurement device, the calibration of the device, the positioning of the wheelset, among others. Ideally, for repeated measurements on the same wheel, different inspection devices should provide similar results. Assessing how good is the agreement between repeated measurements concerns with the measurement error [57]. A previous research study [58] has highlighted the importance of reliable data in the decision-making process surrounding maintenance actions. As mentioned in chapter 2, high reliable products, such as railway wheelsets, are designed to operate without failure for a long time, hence estimating their failure-time distribution is difficult, as only a few units will fail or significantly degrade in a test of practical length at normal use conditions [58]. At the same time, wheelset components are often the main responsible for breakdowns and accidents [4], with wheels being subject to heavy loads caused by wheel-rail contact, generating many different types of defects [59]. Therefore, since the degradation process occurs slowly with usage, small inaccuracies of the measurement device could be reflected in an under- or overestimation of the degradation rates, which could potentially increase maintenance costs over time or increase the likelihood of safety hazards. These facts emphasize the need to accurately measure the parameters, as errors might completely underestimate or overestimate the life of the component under study.

It is a well-accepted principle in engineering that inspection is not perfect, i.e. all measurements have errors [60]. These are simply the difference between the measurements and the true values. Figure 3.3 provides an example of measurement error depiction associated with the deterioration path, illustrated in the form of wheelset wear for two different units/wheelsets. The solid line in the center represents the population average deterioration path, with the individual paths for both units represented in the two dashed lines (above and below). The difference between the centreline (population) and each of the dashed lines (individual) is the random effect, and it represents the individual

heterogeneity that is not explained by the observed covariates. Looking at the lines more closely associated with the individual average paths, it is possible to distinguish the one which represents the actual deterioration path (wear) of the wheelset (thin solid line) and the one which represents the measured (observed) deterioration (thin dotted line). In the former, it is easier to see that there is some autocorrelation between neighbouring points, whereas the latter one is much noisier. The difference between the solid and dotted lines represents the measurement error.

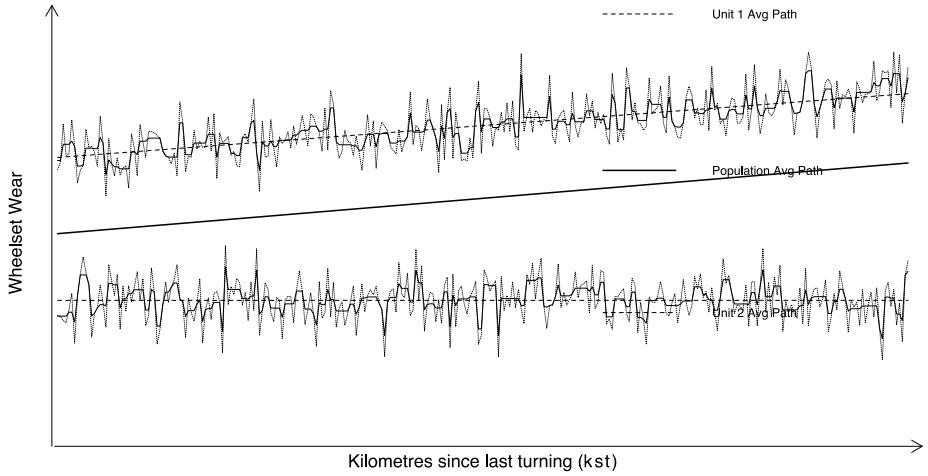

Figure 3.3: Depiction of measurement error. Source: adapted from Yuan and Pandey [61]

As Figure 3.3 suggests, the error associated with repeated measurements for the object of study and a given inspection device can be characterized by two terms: the difference of the true measurement value and the sample mean of the measurement population (bias error or random effect), and the spread of the distribution of the measurements around the sample mean (random or measurement error). In this case study, the true values of the wheel measurements are unknown. In this sense, the use of the word ”error” herein is justified for convention, although when errors are unknown, it may be more appropriate to use the word ”uncertainty” instead.

3.2 Related Work

For paired measurements, the literature offers various methods to assess the agreement of measurements for continuous data. Barnhart et al. [62] organize existing approaches into three main categories: (1) descriptive tools, (2) unscaled summary indices based on absolute differences of measurements (mean squared deviation, limits of agreement, reproducibility variance (ISO), coverage probability (CP), total deviation index (TDI), etc.), and (3) scaled summary indices attaining values between -1 and 1 (correlation coefficients). For a comprehensive review of the main statistical approaches and related concepts, the reader is invited to refer to [62] and [63]. Some other applications in the field are discussed next.

Bland and Altman [64] extend their renowned method, the limits of agreement (LoA), to the more general case when repeated measurements on each of a group of subjects are considered. They highlight two distinct situations for replicated data: the first concerns a series of measurements for the same subject of a quantity that does not vary over the period of observation; the second regards pairs of measurements by two methods of a changing quantity, where it is desired to capture the instantaneous measurement for the subject. The method proposed requires equal replications of each method and each individual for the second case and assumes independence of repeated measurements of a single subject, which is a conservative assumption. Barnhart et al. [65] claim that traditional agreement indices such as intra-class correlation (ICC) or concordance correlation coefficient (CCC) are actually measures of total agreement, and that intra, inter and total method agreement can only be correctly estimated in the presence of replications, so they propose indices for assessing the intra, inter and total agreement based on existence of replications.

Choudhary and Nagaraja [66] present a method for assessing the agreement of two instruments, which also provides information regarding the extent of agreement and nature of disagreement: whether it comes from systematic bias, differences in variability or

correlation. The paired measurements on the same subject are assumed to follow a bivariate normal distribution, and a confidence interval that combines information on the difference in the means, the standard deviation ratios, and the correlation coefficient is constructed using the intersection-union principle. In another work, Choudhary [67] presents a methodology for assessing agreement between two methods of continuous measurements, but this time using a Bayesian mixed model and including a continuous covariate to the response variable. A tolerance band for normally distributed means is constructed, where the mean is modelled using a penalized spline with polynomial basis functions. Choudhary [68] generalizes the previous work for the cases where the repeated measurements are longitudinal or they may be replicates of the same measurement. This application is suitable for a model with covariates, heteroscedasticity and serial correlation in the errors.

In most applications, as the cited examples suggest, measurement methods are compared based on paired (repeated) measurements, i.e. the object of study (usually a homogeneous group of individuals) is measured an equal amount of times with each method and the time between measurements is short enough so that no major changes to the object of study are assumed to occur. Hence, the observed change in the parameter is assumed to come from differences in the methods.

Nevertheless, many of the commonly adopted assumptions may be impractical in some applications: the number of repeated measurements may not be identical across all individuals in the study (unbalanced data), the number of replications of the various methods on the same individual may not be the same (unequal replications), the individuals may not belong to a homogenous group, and the times between measurements may not be short enough to assume no changes in the individuals. For example, Gluer et al. [69] provide a case study where measurement errors for the same technique are higher among osteoporotic patients than in normal subjects and conclude that making inferences for a homogeneous group may be misleading as to determine the technique's precision.

Recent applications deal with the problem of imperfect inspections while relaxing

some of the aforementioned limitations, although they mostly focus on quantifying the measurement uncertainty to improve the maintenance plan and do not assess the agreement between different measurement sources. Dann and Maes [70] study in-line inspections (ILI) of corroded pipelines, where measurement errors may cause the ILI tool to not detect all corrosion features. Usual models for the corrosion growth, which is not directly measurable and needs to be inferred from actual feature sizes measured in two different time epochs, rely on matched features, but the proposed stochastic model does not, although it requires complete sets of data to be efficiently processed and may not achieve the same accuracy of traditional models. Zhang et al. [71] consider a delay time model where imperfect inspections reveal the real states of the system with non-constant probabilities of false positive and false negatives. A three-state system that undergoes a defective state before a failed state is studied, and by analysing the changes in the optimal solutions through several simulations, the authors conclude that maintenance policies that do not incorporate both imperfect maintenance and inspection may result in significant cost loss. Pulcini [72] discusses a perturbed gamma process for describing degradation (and other increasing stochastic processes) in the presence of random measurement errors that are correlated with the actual state of the hidden process. Simulation results revealed that the perturbed gamma process fits data better than the one considering independent measurement errors.

In the specific literature of railway wheelsets, Lin et al. [3] present a Bayesian survival model for locomotive wheels' reliability analysis which can deal with small and incomplete degradation databases, while allowing for the consideration of several covariates. The work is further extended ([73]) using both classical and semi-parametric frameworks to illustrate how degradation data can be modelled and analysed to serve the maintenance decision-making process. Jiang et al. [5] also deal with condition monitoring data for locomotives wheels and highlight that, although remaining useful life (RUL) estimation is a problem commonly tackled in the literature, only a few studies consider measurement errors, i.e.,

they assume inspection measurements are perfect and without error. They develop a wear model that considers both stochastic degradation and measurement error based on a discrete state space model to estimate the RUL. A case study is used to illustrate the method showing that it can improve the accuracy of the estimated RUL in the long run.

In a recent application, Urda et al. [74] present a scaled dynamometric wheelset equipped with two distinct technologies: strain gauges and distance lasers that measure wheel web deflection. Measurements of the two technologies (installed on the same scaled vehicle) are compared in terms of ability to measure wheel-rail contact forces and other factors, such as ease of use, costs and possible errors introduced by setting up the piece of equipment. However, the comparison is based on having the scaled system ready, which may not be easily reproducible in other studies. Huang et al. [75] present an experiment involving bearing signals from a test rig in order to simulate an axle box with faulty bearing supporting the wheelset and they make the consideration of measurement noises in signal form, although no comparison with other methods was performed. The proposed method, the fast extended singular value decomposition package (FESVDP), extracts information from the signal in different resonance bands and determine the faulty pattern. Bosso et al. [76] also use axle box measurements, focusing on the problem of detecting wheel flats by using the vertical acceleration measurements and time-domain analysis. They run experimental tests with an actual operating vehicle and conclude that for the application of the proposed method in real operations it is necessary to perform a specific calibration of the algorithm to adapt it to the particular vehicle and line being used. Finally, Alemi et al. [77, 78] work with wheel measurements coming from multiple sensors to feed a ”fusion” method which reconstructs a new signal containing the pattern of the contact force that is a function of the wheel defect. They show the efficacy of the method to detect wheel defects, however, when the contact force ratio is not relatively large, the measurement noise decreases the similarity of the reconstructed signal, decreasing the ability of the method to detect the minor defects. Although all these works deal with

railway wheel measurements and draw some important conclusions regarding how the errors impact wheels and estimation of defects or RUL, none of them present analysis comparing different sources of data.

Assessing agreement is very rarely done using operation or condition monitoring data by contrasting several devices in a real-world case study, i.e. using inspection data from a system in operation, which is the goal of this study. As mentioned before, inspection data may bring additional challenges to the problem of assessing the agreement between inspection devices, as data may be unbalanced, with unequal replications, the individuals may not belong to a homogenous group, and deterioration data may display serial correlation. The method adopted in this study is a mixed-model framework, which provides a systematic way to account for between- and within-subject variability and for the serial correlation among measurements. Some related work using LMM has been published in the literature. Their main ideas and the differences related to the present work will be discussed next.

Andrade and Stow [52] discuss LMM and generalized LMM to model the wear trajectories of railway wheelsets in terms of the evolution of the wheel flange thickness, the flange height and the tread diameter. They also incorporate the statistical modeling of damage trajectories by considering the probabilities of various types of wheel tread damage, such as rolling contact fatigue (RCF), wheel flats and cavities. In a related work, Andrade and Stow [53] present an application of LMM and SFA (stochastic frontier analysis) in the study of the variability of wheelset turning operation. The authors show that the error component associated with the inefficiency of the under-floor wheel lathe operators dominates the variability around the mean of the diameter loss due to turning, although they only consider inspection data coming from one source (the under-floor wheel lathe). They conclude that there are significant differences among operators that are worth being further investigated, in terms of understanding the impact of different turning ”approaches” in the wheels’ life cycle. Another work by Andrade and Stow [54]

studies the potential cost savings of introducing 'Economic Tyre Turning' (ETT) in the maintenance practice of British train operating companies. They specify several statistical models to predict the evolution of wheel wear and damage, and then they use a Monte Carlo simulation procedure to assess the different models, while also controlling for random effects associated with train unit, individual vehicles within units and measurement month, showing that the consideration of those random effects is important in the deterioration path estimation. Life cycle costs for different maintenance strategies are presented and compared, leading to the conclusion that ETT can provide substantial savings. In a different context, Yuan and Pandey [61] use a nonlinear mixed-model approach for modelling and predicting degradation in a nuclear piping system with data obtained from periodic inspection. They present a case study showing that a mixed-effects model is an effective approach for modelling degradation data of engineering systems.

3.2.1 Main Contributions

The main contribution of this chapter is to use inspection data to develop a model that can successfully explain how the wheelset parameters change over some distance measure (e.g. kilometers since last turning/renewal), while controlling for some other covariates and, based on this model, assess the agreement between the three different types of measurements (manual, laser and wheel lathe). The assessment of agreement is particularly challenging when real-world inspection data is considered, since data is often serially correlated, unbalanced and unequally replicated. In light of the previous section, although there are studies in the literature which consider mixed models and inspection data, they do not explicitly deal with the assessment of the agreement. The present work aims to fill this research gap.

3.3 Preliminares

This section is devoted to preliminaries on linear and linear-mixed models, LMs and LMMs, respectively.

3.3.1 Linear-Mixed Models

The option for using LMMs comes from the fact that they are flexible, albeit simple method to account for known differences in the variance of the different groups. Another motivation to use LMMs is that the response variable, wheel wear, depends on the characteristics of the rail vehicle [79]. By treating the effect of a variable as random (opposed to fixed), the interest shifts from knowing the performance of the various levels of that variable presented on the dataset to knowing the variation among all levels present in the population. A more intuitive example would be to consider the variable kilometers since last turning (kst) in the case study: instead of one single "slope" for the variable kst (fixed effects), the random effects approach allows for random intercepts (i.e., different baseline response values), where the amount of variation in the average response caused by a given kst is estimated (random effect) and added to the curve. Moreover, as highlighted by Yuan and Pandey [61], traditional regression models are inadequate for modeling periodic inspection data, as they may not capture potential correlation among repeated measurements in a structured way. Furthermore, the nature of the inspection data, as highly unbalanced, makes it not appropriate for being analyzed using only fixed effects regression techniques.

In addition to allowing the correlation among observations to be taken into account, another advantage of the LMM approach is an effective partition of the overall variation of the dependent variable into components corresponding to different levels of data hierarchy [80], i.e. random effects associated with different groups. This is especially useful for the purpose of this study.

Given deterioration data in the form of wheelset wear measured by either ΔFt, ΔFh

and ΔqR, i.e., (y_i, kst_i), $i = 1, \cdots, n$ where y_i is the measured value (deterioration) for the parameter being analyzed after kst_i (kilometers since last turning), a simple linear model (LM) is expressed as:

$$y_i = \beta_0 + \beta_1 kst_i + \varepsilon_i, \qquad \varepsilon \sim N(0, \sigma^2) \tag{3.1}$$

In the above model, the intercept β_0 can be interpreted as the post-turning wheel parameter (i.e., either Ft, Fh or qR) and β_1 as an average wear rate. The response y_i is the variation/wear (ΔFt, ΔFh and ΔqR) as a function of the post-turning wheel parameter and the continuous explanatory variable kilometers since last turning kst. The random noises $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are assumed independent and identically distributed (i.i.d.) normal random variables with mean zero and constant variance σ^2. The linear model predicts the degradation as a function of the kilometers since last turning. The prediction of a new observation $\hat{y}_{pred}$ corresponding to $kst = kst_0$ is obtained by substituting $kst = kst_0$ in the fitted regression model, as follows:

$$\hat{y}_{pred} = \hat{\beta}_0 + \hat{\beta}_1 kst_0 \tag{3.2}$$

where $\hat{\boldsymbol{\beta}} = \left(\hat{\beta}_0, \hat{\beta}_1\right)^T = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$ is the usual ordinary least squares (OLS) solution of the linear regression model which minimizes the residual sum of squares (RSS) defined, in matrix form, by RSS $= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$, where $\boldsymbol{X}$ is the $n \times p$ model matrix (p corresponding to the dimension of $\boldsymbol{\beta}$).

An unbiased estimator for the variance is given by $\hat{\sigma}^2 = \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right)/(n-p)$. Under the linear model assumptions of independence and normality, it holds that the OLS estimator $\hat{\boldsymbol{\beta}}$ follows a multivariate normal distribution, i.e., $\hat{\boldsymbol{\beta}} \sim N\left(\beta, \sigma^2(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)$. Therefore, the variance of $\hat{y}_{pred}$ corresponding to $kst = kst_0$ or in matrix form $\mathbf{x}(kst_0) = (1, kst_0)$, includes both the variance of the random noise ε and the uncertainty associated

with the parameters, as follows:

$$\text{Var}(\hat{y}_{pred}) = \text{Var}\left(\hat{\beta}_0 + \hat{\beta}_1 kst_0\right) + \text{Var}(\varepsilon) = \left(\mathbf{x}(kst_0)(\boldsymbol{X}^T\boldsymbol{X})^{-1}\mathbf{x}^T(kst_0) + 1\right)\sigma^2 \quad (3.3)$$

The result in Equation 3.3 considers that the observed future value will fluctuate around the mean response at $kst = kst_0$, which is the reason why the random noise variance or, in other words, the variance of an individual observation, is added.

As highlighted by Yuan and Pandey [61], unlike the case of a single measurement for each unit, repeated measurements (several times at different instances) will no longer be independent. The linear regression model presented above has its limitations to model the dependence among the repeated measurements. In some cases, a covariance matrix $\sigma^2\boldsymbol{\Sigma}$ suffices to model the dependence, however, estimation of $\boldsymbol{\Sigma}$ may not be easy, especially in the context of degradation data which is commonly unbalanced.

A linear mixed model (LMM) offers a more structured way to account for the correlation among the measurements. Cluster correlation arises naturally when measuring the same units repeatedly [81] and the observed (measured) values are not independent. For cases like that, an extension of the linear model, where the covariance structure of the error terms can be specified, is given by [81]:

$$\boldsymbol{y}_i = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{Z_i}\boldsymbol{b_i} + \boldsymbol{\varepsilon_i} \quad (3.4)$$

Equation 3.4 considers a single level of grouping, with N groups indexed by $i = 1, \cdots, N$, each containing n_i observations, where $\boldsymbol{y_i}$ is the $n_i \times 1$ vector of responses, $\boldsymbol{X_i}$ is an $n_i \times p$ known, possibly non-full-rank matrix of fixed predictors and $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown fixed parameters. The $\boldsymbol{Z_i}$'s are known $n_i \times q$ full-rank matrices used to specify membership in the various subgroups and the $\boldsymbol{b_i}$'s are $q \times 1$ random vectors, $\boldsymbol{b_i} = (\theta_{i1}, \cdots, \theta_{iq})$, such that $\text{E}(\boldsymbol{b_i}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{b_i}) = \boldsymbol{D}$. It is further assumed that the

$n_i \times 1$ random vector $\boldsymbol{\varepsilon_i}$ is normally distributed with $\mathrm{E}(\boldsymbol{\varepsilon_i}) = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon_i}) = \sigma^2 \boldsymbol{I}_{n_i}$, i.e., the residual errors are assumed independent of each other just like in the LM. It is also considered that $\mathrm{cov}(\boldsymbol{b_i}, \boldsymbol{b_j}) = \mathbf{0}$ for $i \neq j$, and $\mathrm{cov}(\boldsymbol{b_i}, \boldsymbol{\varepsilon_i}) = 0$ for all i, i.e., the vectors of random effects and residual errors for different groups are independent of each other. Under those assumptions, degradation for a unit i is normally distributed with $\mathrm{E}[\boldsymbol{y_i}] = \boldsymbol{X_i}\boldsymbol{\beta}$ and $\mathrm{cov}(\boldsymbol{y_i}) = \boldsymbol{V_i} = \boldsymbol{Z_i}\boldsymbol{D}\boldsymbol{Z_i}^T + \sigma^2 \boldsymbol{I}_{n_i}$, i.e., the z's only enter in the covariance structure and the mean is only influenced by the fixed predictors x's. For the general specification of the model for multiple grouping levels, see [81].

The $\boldsymbol{Z_i}\boldsymbol{b_i}$ terms in the LMM allows for modeling random effects, or the individual heterogeneity that is not explained by the observed covariates. Although $\mathrm{cov}(\boldsymbol{\varepsilon_i}) = \sigma^2 \boldsymbol{I}_{n_i}$ implies independent errors, the observed values of response variable for the same unit (repeated measurements) are no longer independent, as $\mathrm{cov}(y_{ij}, y_{ik}) = \boldsymbol{z_{ij}}\boldsymbol{D}\boldsymbol{z_{ik}^T} > 0$ for $j \neq k$, which is an important characteristic of the mixed-effects model.

The maximum likelihood method is usually adopted to find estimates for the LMM model from the marginal distribution $\boldsymbol{y_i} \sim N_{n_i}\left(\boldsymbol{X_i}\boldsymbol{\beta}, \boldsymbol{Z_i}\boldsymbol{D}\boldsymbol{Z_i^T} + \sigma^2 \boldsymbol{I_{n_i}}\right)$. The restricted maximum likelihood (REML) method is also usually employed as the ML estimates tend to underestimate the variance components, although the methods should give approximately the same solution when the number of parameters is small [61]. The log-likelihood for unit i can be obtained as follows:

$$\ell_i(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) = -\frac{1}{2}\left(\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta}\right)^T \boldsymbol{V_i}^{-1}\left(\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta}\right) - \frac{1}{2}\ell_n|\boldsymbol{V_i}| \tag{3.5}$$

Estimates of $\boldsymbol{\beta}$, σ^2 and $\boldsymbol{\theta}$ are usually obtained using a log-profile-likelihood for the variance components [81], which results from plugging into Equation 3.5 the weighted least squares estimator for $\boldsymbol{\beta}$, namely $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \left(\sum_{i=1}^{n}(\boldsymbol{X_i}^T\boldsymbol{V_i}^{-1}\boldsymbol{X_i})\right)^{-1}\left(\sum_{i=1}^{n}(\boldsymbol{X_i}^T\boldsymbol{V_i}^{-1}\boldsymbol{y_i})\right)$. Then $\hat{\sigma}^2$ and $\hat{\boldsymbol{\theta}}$ are obtained by maximizing the log-profile-likelihood as in [81]. For a specific unit, the estimate for $\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} = \boldsymbol{D}\boldsymbol{Z_i}^T\boldsymbol{V_i}^{-1}\left(\boldsymbol{y_i} - \boldsymbol{X_i}\hat{\boldsymbol{\beta}}\right)$. For a unit with measurements $\boldsymbol{y_i}$, prediction of wear at $kst = kst_0$, considering the two covariate vectors at $\mathbf{x}_i(kst_0)$ and

$\mathbf{z}_i(kst_0)\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}$, is given by:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i,pred}} = \mathbf{x}_i(kst_0) + \hat{\boldsymbol{\beta}} + \mathbf{z}_i(kst_0)\hat{\boldsymbol{\theta}}_{\boldsymbol{i}} \tag{3.6}$$

The fitted values, after some algebraic manipulations, can be obtained as:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \left(\sigma^2 \boldsymbol{V_i}^{-1}\right) \boldsymbol{X_i}\hat{\boldsymbol{\beta}} + \left(\boldsymbol{I}_{n_i} - \hat{\sigma}^2 \boldsymbol{V_i}^{-1}\right) \boldsymbol{y_i} \tag{3.7}$$

Equation 3.7 shows that the computation of the fitted values involves two terms, and can be interpreted as a weighted average of the population mean path $\boldsymbol{X_i}\hat{\boldsymbol{\beta}}$ and the observed data $\boldsymbol{y_i}$, with denominator $\boldsymbol{V_i}$. In practice, if within-unit variability $\hat{\sigma}^2$ is large in comparison to the cross-unit variability $\boldsymbol{D}$, then more weight is assigned to the first term, whereas the opposite also holds true. By separating the variability into these two parts, the LMM is able to better quantify the sampling uncertainty of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ [61].

In the next section, a case study on the change in the flange height parameter (ΔFh) is performed to illustrate differences between LM and LMM in dealing with the assessment of agreement between different inspection devices. Then, a summary of similar analyzes is presented for the change in flange thickness (ΔFt) and in flange slope (ΔqR) parameters.

3.4 Assessing Agreement Between Wheelset Inspection Devices

3.4.1 Exploratory Data Analysis

The following inspection data was acquired from a Portuguese railway company that transports passengers in a single line, and which extends 54 kilometers and serves 14 stations. A 17-year interval database, ranging from January 2001 up to July 2019, with measurements coming from three different inspection devices (gauge device, laser device and under-floor wheel lathe) is considered. The company operates 18 Electrical Multiple Units (EMUs). Each unit has 4 cars, and each car has eight wheels (i.e. four wheelsets). Figure 3.4 provides a schematic representation of a four-car unit.

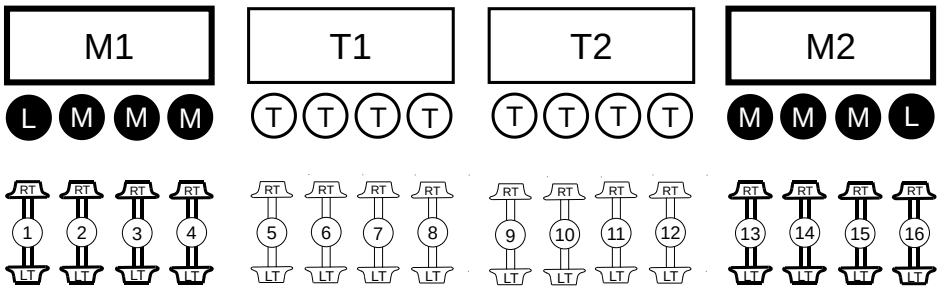

Figure 3.4: Schematic representation of a 4-car unit

Figure 3.4 depicts one EMU with four vehicles, where each vehicle has eight wheels (i.e., four wheelsets). The EMU has 16 wheelsets, which are divided into three categories according to their relative position, as shown in Figure 3.4, where position 'L' represents a leader motor wheelset, 'T' represents a trailer wheelset and 'M' stands for a simple motor wheelset. Each wheelset has two sides: left ('LT') and right ('RT').

The database contained the following information: unit number, unit running kilometres (cumulative kilometers), measurement type (manual, laser or turning), date, wheelset position (1 up to 16), side of wheel (left or right), flange thickness (Ft), flange height (Fh) and flange slope (qR).

From this original database, it was possible to compute other variables of interest: kst (kilometers since last turning), ΔFh (change in flange height since last turning due to wear), ΔFt (change in flange thickness since last turning due to wear), ΔqR (change in flange slope since last turning due to wear). Table 3.1 provides an overview of the aforementioned variables as well as some relevant descriptive statistics.

Table 3.1: Main variables and descriptive statistics (mean, standard deviation, minimum and maximum)

Variable	Description	Type	Mean	Std	Min	Max
ΔFh	Change in the flange thickness (in mm) since last turning due to wear	Continuous	-0.07	2.00	-8.80	7.10
ΔFt	Change in the flange height (in mm) since last turning due to wear	Continuous	1.56	1.44	-7.90	9.80
ΔqR	Change in the flange slope (in mm) since last turning due to wear	Continuous	-0.20	1.89	-6.60	5.30
kst	Kilometers since last turning	Continuous	110814	79551	133	343662
M	Measurement Type (Manual, Laser or Under-floor Wheel Lathe)	Nominal	-	-	-	-
U	Unit Number (1 up to 36)	Nominal	-	-	-	-
W	Wheelset Position (L, M or T)	Nominal	-	-	-	-
S	Side of the wheel (L or R)	Nominal	-	-	-	-

As mentioned earlier, date ranges for measurements did not coincide, with the following

ranges being available: manual measurements ranging from July 2001 to March 2015, laser measurements from February 2017 to July 2019 and turning measurements from October 2000 to July 2019. Manual measurements gradually stopped being performed when the laser device was acquired, although it was never the case that a manual measurement was immediately followed by a laser measurement, and hence direct (paired) comparisons could not be performed. Furthermore, not all turning measurements were available: in many cases, an interval of over a year would separate two turning measurements, something that, in practice, should not occur. Hence, it was decided not to include those in the final database, to avoid potential bias on the results.

Under these conditions, the following exploratory analysis is performed. Figure 3.5 shows the histograms of the measurements for the main dependent variables ΔFt, ΔFh, ΔqR.

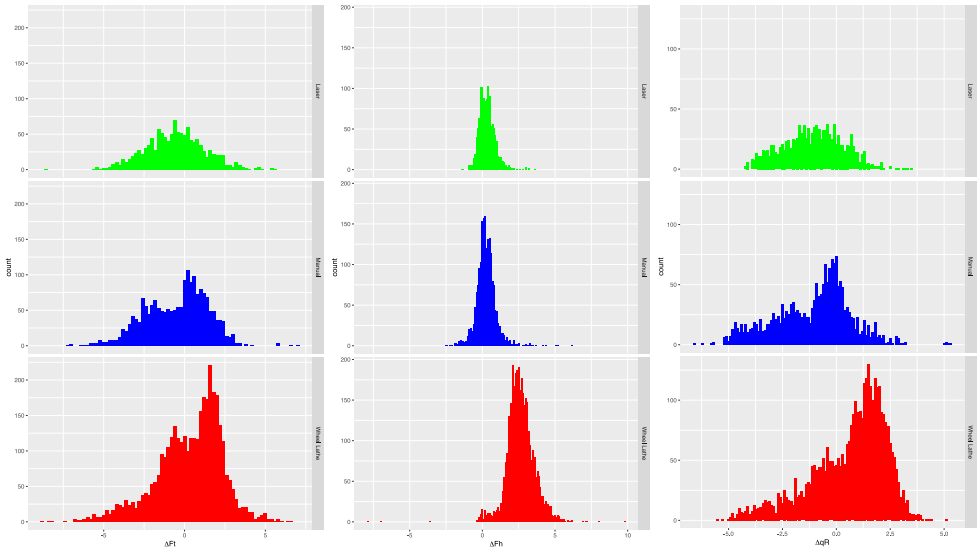

Figure 3.5: ΔFt, ΔFh, ΔqR histograms per type of measurement

The histograms suggest that the distribution of the different statistics, when stratified by the three types of measurements, are not the same. The means for the wheel lathe measurements distributions seem to be slightly shifted in comparison to the other measurements distributions, which could be explained by the fact that wheels are only measured in the wheel lathe when kst is above some threshold (for inspection) or it

is believed that the wheel might be a candidate for turning (due to wear or damage). Bimodality in some of the sampling distributions also reveals that some factors (e.g. wheelset position, side of the wheel, etc.) might be influencing the measurements. This suggests that the linear model should account for possible differences in the means. This initial exploratory data analysis provides a better understanding of the data. However, inferences regarding the predominance of one type of measurement over the others in terms of precision are still not completely understood. Therefore, a more systematic approach is considered in the following subsections on each dependent variable ΔFh, ΔFt, ΔqR, respectively.

3.4.2 Change in Flange Height Due to Wear ΔFh

The case study focuses on the deterioration of wheels observed as the change in the flange height due to wear (ΔFh). The other two parameters, i.e. the change in the flange thickness due to wear (ΔFt) and the change in the flange slope due to wear (ΔqR), will not be described in the same level of detail, as the analysis is very similar. Instead, the main results will be briefly discussed in their corresponding subsections.

The choice of flange height as the main parameter comes from the fact that the wheels are always reprofiled to a same initial flange height, i.e., the wheel-lathe machine always brings the wheel flange height back to a same pre-established value, which does not happen for the other two parameters. In this sense, the post-turning wheel parameter value does not contain the bias originated from having different machining set-ups. Another aspect of the analysis is that it considers only leader wheelsets, i.e., wheelset type 'L', which is known to deteriorate faster than the types 'M' (motor) and 'T' (trailer). In practice, this means that only two wheelsets (out of 16 as shown in Figure 3.4) will be considered per EMU, so that the variability caused by different wheelset positioning does not interfere with the analysis. Other known factors affecting the wear rates considered in this study are kst (kilometers since last turning) and unit number (U). The variable wheel side (S) was not statistically

significant in any model, hence, this variable will be omitted from the analysis.

One of the characteristics of this study, as mentioned earlier, is that measurements were made at different times (not paired) and the number of repeated measurements is not the same across different units or inspection devices (unbalanced). Table 3.2 shows the number of measurements per unit and per measurement type.

Table 3.2: Number of measurements per train unit number and measurement type (laser, manual and wheel lathe)

Unit Number	Laser	Manual	Wheel Lathe	Total
3501	4	4	12	20
3502	2	6	10	18
3503	4	6	12	22
3504	2	8	12	22
3505	2	8	12	22
3506	6	8	14	28
3507	6	6	16	28
3508	4	4	10	18
3509	6	8	8	22
3510	4	2	14	20
3511	4	6	10	20
3512	2	12	14	28
3513	16	8	12	36
3514	2	4	14	20
3515	4	4	12	20
3516	8	6	10	24
3517	4	10	16	30
3518	8	6	10	24
3551	6	4	10	20
3552	4	8	10	22
3553	4	6	16	26
3554	2	8	12	22
3555	0	8	12	20
3556	4	8	14	26
3557	2	4	14	20
3558	6	8	12	26
3559	6	8	10	24
3560	4	2	16	22
3561	4	6	10	20
3562	0	12	13	25
3563	10	6	10	26
3564	2	4	12	18
3565	4	4	10	18
3566	6	8	10	24
3567	2	10	14	26
3568	2	10	8	20
Total	156	240	431	827

A simple LM (Equation 3.1) is fitted first, where the response y_i is the wear ΔFh and the continuous explanatory variable is kilometers since last turning (kst) multiplied by 1000 (i.e., the units come in the form of thousand kilometers: k km). In this model, the intercept β_0 can be interpreted as the post-turning wheel parameter Fh and β_1 as an average wear rate. The measurements are assumed independent of each other. The model has an

adjusted R^2 of 0.529 and a $F_{1,825}$ statistic of 926.5 and, hence, it has a corresponding p-value of less than 2.2×10^{-16}, making it statistically significant. The least-squares estimates of the parameters (estimate and standard error inside parentheses) are:

$$\hat{\beta}_0 = -0.1823(0.0645), \qquad \hat{\beta}_1 = 0.0150(0.0005), \qquad \hat{\sigma} = 1.081$$

The linear model residuals stratified by unit number and measurement type are displayed in Figure 3.6.

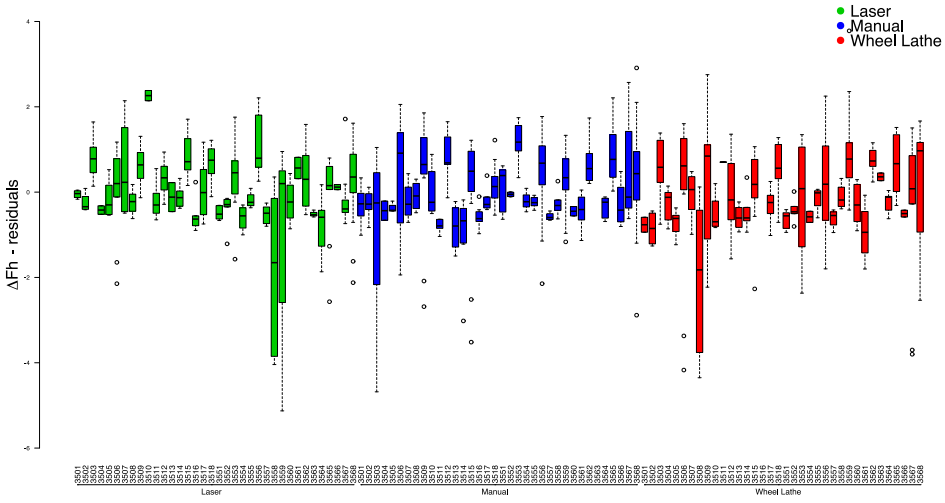

Figure 3.6: ΔFh - Linear model residuals per unit number and type of measurement

Figure 3.6 presents all the 36 units boxplots of residuals by each type of measurement, in ascending order by unit number, which makes the comparison of unit by unit across measurement types possible. The boxplots suggest the existence of significant differences among units. Some units, for example, display all negative residuals, indicating potential bias. However, because of the individual biases, it is very complicated to compare the different types of measurements. If individual linear models are fitted for each unit, it is easier to notice the existence of a unit effect, motivating the LMM approach. Figure 3.7 shows the 95% confidence intervals for the coefficients of initial wear ($\hat{\beta}_0$) and wear rate ($\hat{\beta}_1$) of the linear models by unit number. The 95% confidence intervals ($CI_{95\%}$) for the

coefficients of the estimated linear model are $[0.0557, 0.3088]$ for the initial wear ($\hat{\beta}_0$), and $[0.0140, 0.0159]$ for the wear rate ($\hat{\beta}_1$). If the individual intervals in Figure 3.7 are compared to these main linear model intervals, many of the former ones are not covered in the main $CI_{95\%}$, putting into evidence that an LMM approach might be necessary to account for the individual differences.

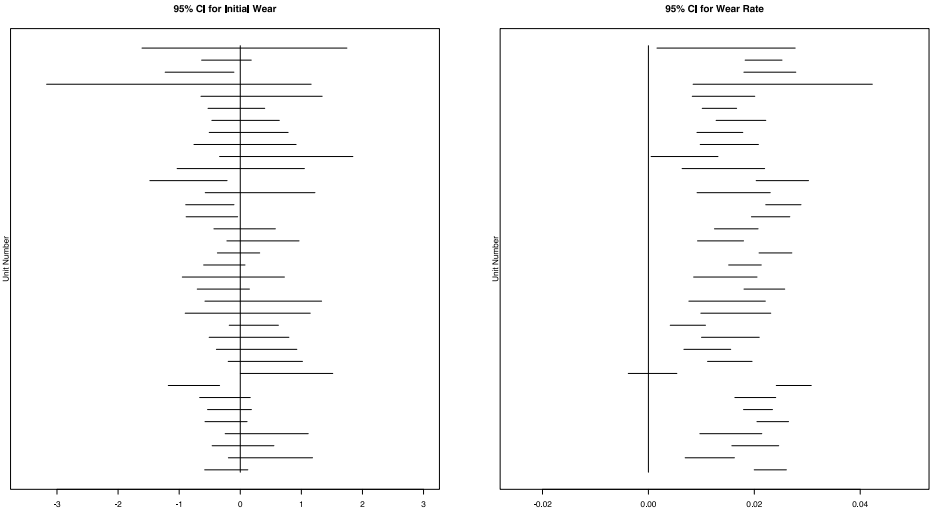

Figure 3.7: 95% Confidence Interval (CI) for the coefficients of initial wear ($\hat{\beta}_0$) and wear rate ($\hat{\beta}_1$) of the linear models by unit number

The LM analysis above suggests the need for an LMM approach, in which it is assumed an overall mean and the fixed effect of kilometers since last turning (kst), and it also includes the variables kst and unit number U as random effects. This LMM is equivalent to, for a given unit i and its j^{th} measurement/observation:

$$y_{ij} = \beta_0 + \beta_1(kst)_{ij} + \theta_{unit_i} + \theta_{kst_{ij}} + \varepsilon_{ij} \tag{3.8}$$

In the above model $\theta_{unit_i} \sim N(0, \sigma^2_{unit})$, $\theta_{kst_{ij}} \sim N(0, \sigma^2_{kst})$. The REML criterion associated with this model is 2236.6 and the REML estimates of the parameters (estimate and standard error inside parentheses) are:

$$\hat{\beta}_0 = 0.1471(0.0932),\ \hat{\beta}_1 = 0.0150(0.0006),\ \hat{\sigma}_\varepsilon = 0.6863,\ \hat{\sigma}_{unit} = 0.2201,\ \hat{\sigma}_{kst} = 0.8025$$

The estimate for the overall mean $\hat{\beta}_0$ changes significantly from the LM ($\hat{\beta}_0 = -0.1823$ (0.0645)) to the LMM ($\hat{\beta}_0 = 0.1471$ (0.0932)), which does not occur for $\hat{\beta}_1$, that remains about the same. However, the estimated standard error for the residuals drops significantly from 1.081 in the LM to 0.6863 in the LMM, showing that ignoring the differences among different groups can potentially cause one to incorrectly interpret errors coming from the group structure of the data as measurement errors coming from the inspection devices.

A similar plot to Figure 3.6 presenting all the 36 units boxplots of residuals by each type of measurement is displayed for the residuals of the LMM in Figure 3.8:

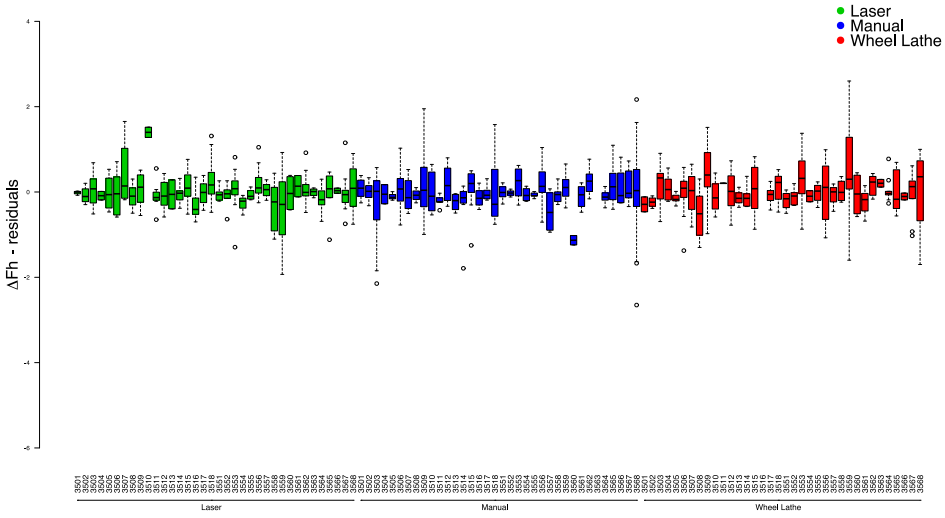

Figure 3.8: ΔFh - Linear model residuals per unit number and type of measurement

As expected, the boxplots in Figure 3.8, in general, have alternate signs and are smaller (when compared with Figure 3.6), showing the effectiveness of the LMM in handling the group structure of the data. Comparing the fit of both models in terms of Root Squared Mean Error (RMSE) is also useful to see how much better the LMM fits the data. Figure 3.9 shows the comparison with the RMSE for each model:

As mentioned earlier, the variability in the response ΔFh is explored through an LMM approach. Since true values for the measurements are unknown, it is assumed that the chosen model can accurately represent the association between the response variable and kilometers since last turning (kst), after controlling for the variation within unit number and

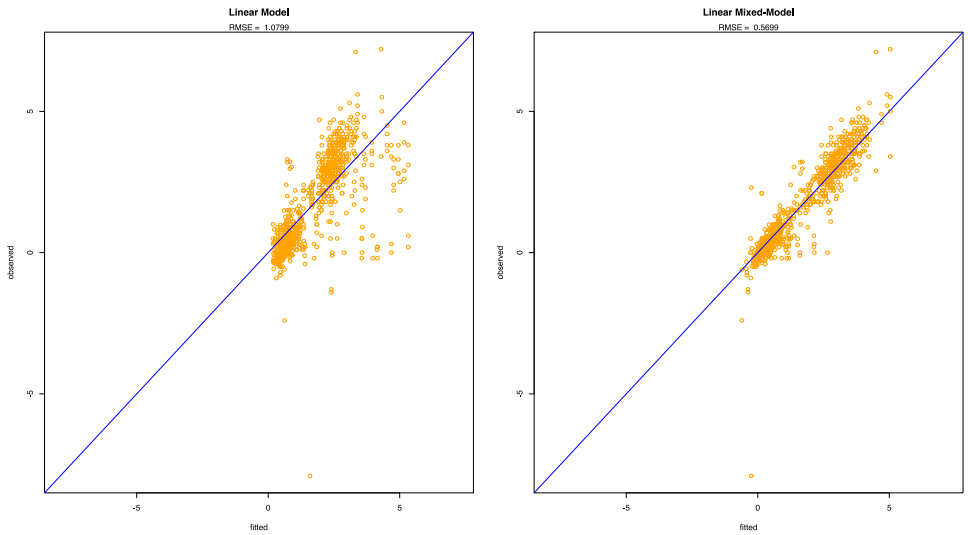

Figure 3.9: ΔFh - Comparison of fitted values for LM and LMM models

kst. By assuming the model is true, looking at the residuals per measurement type allows a simple way to compare them, and provide an answer to the research question, on what is the performance of each inspection device in terms of uncertainty of the measurement error. Therefore, Figure 3.10 plots the residual per measurement type:

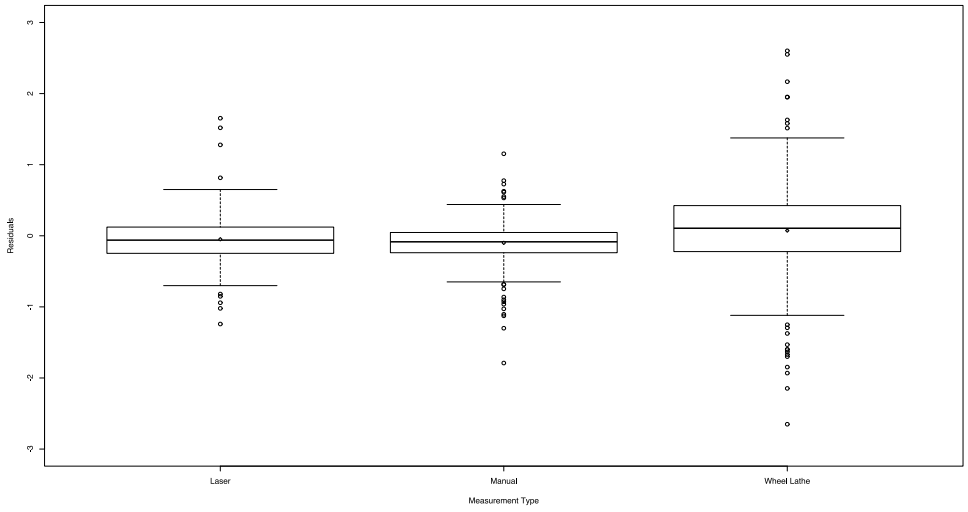

Figure 3.10: ΔFh - Analysis of LMM residuals per measurement type

Following the goal of the study, a Levene's test can be used to test the hypothesis that the three measurement types have homogeneous variances. The test statistic has an

approximate $F_{k-1,N-k}$ distribution with degrees of freedom equal to the number of groups (k) minus 1, and sample size (N) minus number of groups (k). The obtained result for the $F_{2,824}$ statistic was 22.886 with a corresponding p-value of 2.124×10^{-10}, hence, the null hypothesis is rejected, as the residuals for different categories show no strong evidence that there is homogeneity of variances across groups, i.e. when the underlying LMM is assumed to hold, there seems to exist differences in the spread of the residuals according to the measurement type. Testing each pair of measurement type separately, the null hypothesis is rejected for all pairs at a significance level of 0.05, as shown in Table 3.3:

Table 3.3: Levene's statistical test results for each measurement type pair

Pair	$F_{k-1,N-k}$	p-value	Conclusion
Laser *vs.* Manual	4.315	0.0385	Reject null hypothesis
Laser *vs.* Wheel Lathe	13.338	0.0002	Reject null hypothesis
Manual *vs.* Wheel Lathe	35.716	3.712×10^{-9}	Reject null hypothesis

Table 3.3 results indicate that there is no evidence to support the homoscedasticity hypothesis - equal variances among different measurement types, i.e., measurement types have statistically significant different variances. At a lower significance level, e.g. $\alpha = 0.01$, one does not reject the hypothesis that the variances of laser and manual measurements are equal, which can also be inferred by visually inspecting the boxplots in Figure 3.10. In this sense, special attention should be given to understanding the spread observed in wheel lathe measurements, before understanding differences between manual and laser inspections.

3.4.3 Change in Flange Thickness Due to Wear ΔFt

A similar analysis can be conducted for the change in flange thickness due to wear, ΔFt. The results in terms of LM and LMM and the Levene's test for homogeneity of variances will be briefly discussed.

The LM has an adjusted R^2 of 0.099 and a $F_{1,825}$ statistic of 91.45 and, hence, it has a corresponding p-value of less than 2.2×10^{-16}, making it statistically significant. The model

also has a corresponding RMSE of 2.0551. The least-squares estimates of the parameters (estimate and standard error inside parentheses) are:

$$\hat{\beta}_0 = -1.4235(0.1227), \qquad \hat{\beta}_1 = 0.0089(0.0009), \qquad \hat{\sigma} = 2.0580$$

The LMM, which was built in the same framework considered earlier for the flange height model, has an associated RMSE of 1.0741, showing a significant improvement if compared to the LM. The estimated standard error for the residuals also drops significantly from 2.0580 in the LM to 1.2975 in the LMM and the REML criterion at convergence is 3306.3. Finally, the model estimates are:

$$\hat{\beta}_0 = -1.378(0.1765),\ \hat{\beta}_1 = 0.0078(0.0013),\ \hat{\sigma}_{\varepsilon} = 1.2975,\ \hat{\sigma}_{unit} = 0.3176,\ \hat{\sigma}_{kst} = 1.6016$$

By assuming the above LMM model is true, the residuals per measurement type plot are displayed in Figure 3.11 below:

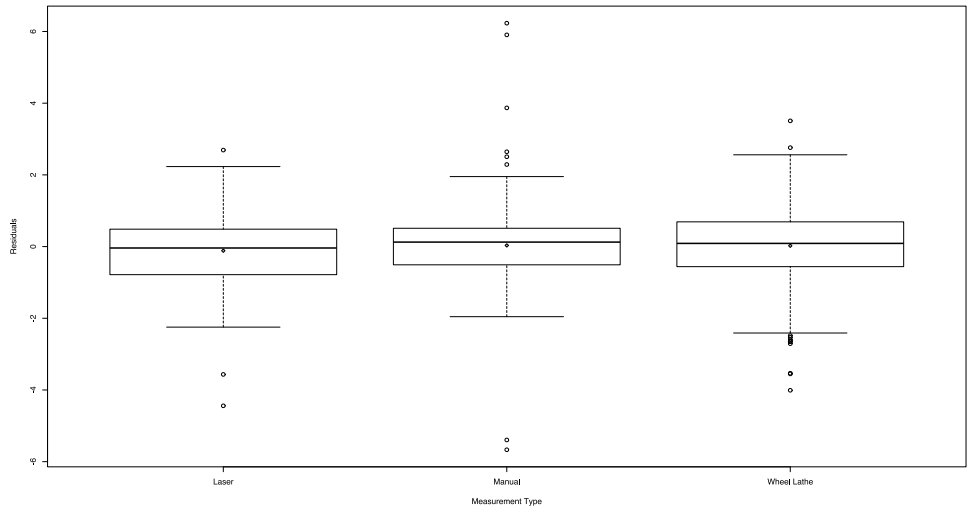

Figure 3.11: ΔFt - Analysis of LMM residuals per measurement type

If the Levene's test is run, the obtained result for the $F_{2,824}$ statistic is 1.614 with a corresponding p-value of 0.1997, hence, the null hypothesis of homogeneity of variances cannot be rejected at a significance level of 0.05, i.e., there is not statistically significant

evidence that the inspection devices are different for measuring the flange thickness parameter.

3.4.4 Change in Flange Slope Due to Wear ΔqR

The last analysis is on the change in flange slope due to wear, ΔqR. The LM has an adjusted R^2 of 0.2066 and a $F_{1,825}$ statistic of 216.1 and, hence, it has a corresponding p-value of less than 2.2×10^{-16}, making it statistically significant with a corresponding RMSE of 1.7578. The least-squares estimates of the parameters (estimate and standard error inside parentheses) are:

$$\hat{\beta}_0 = -1.8437(0.1049), \qquad \hat{\beta}_1 = 0.0117(0.0008), \qquad \hat{\sigma} = 1.76$$

The LMM, which was built in the same framework considered earlier for the flange height and thickness models, has an associated RMSE of 0.8379, showing a significant improvement if compared to the LM, as seen on previous analyses. The estimated standard error for the residuals also drops significantly from 2.0580 in the LM to 1.4367 in the LMM and the REML criterion at convergence is 2973.5. Finally, the model estimates are:

$$\hat{\beta}_0 = -1.8644(0.1501),\ \hat{\beta}_1 = 0.0109(0.0012),\ \hat{\sigma}_{\varepsilon} = 1.0219,\ \hat{\sigma}_{unit} = 0.1922,\ \hat{\sigma}_{kst} = 1.4367$$

By assuming the above LMM model is true, the residuals per measurement type plot are displayed in Figure 3.12 below:

The obtained Levene's statistical test result for the $F_{2,824}$ statistic is 3.9151 with a corresponding p-value of 0.02031, hence, the null hypothesis is rejected, as the residuals for different categories show no strong evidence that there is homogeneity of variances across groups. Testing each pair separately under a significance level of 0.05, the null hypothesis is rejected only for the pair Manual *vs.* Wheel Lathe, as shown in Table 3.4:

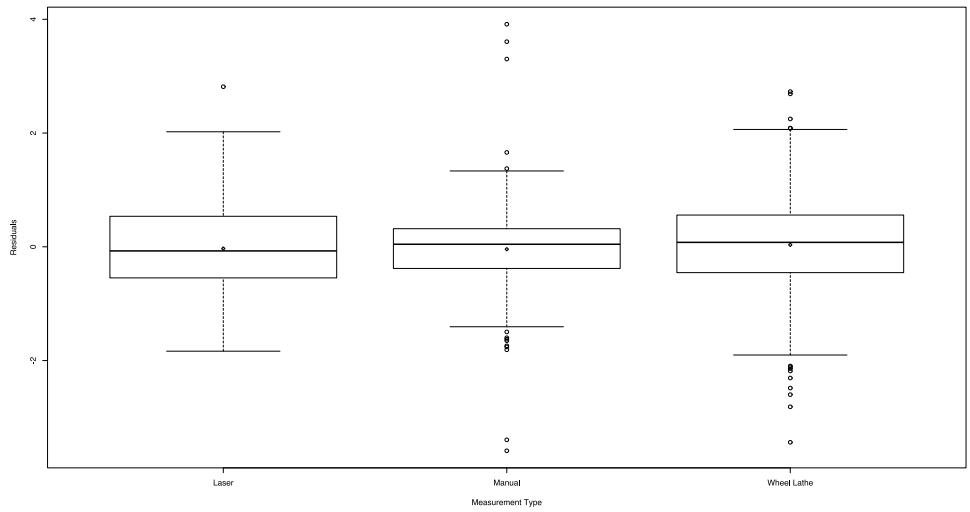

Figure 3.12: ΔqR - Analysis of LMM residuals per measurement type

Table 3.4: Levene's statistical test results for each measurement type pair

Pair	$F_{k-1,N-k}$	p-value	Conclusion
Laser *vs.* Manual	2.569	0.1098	Do not reject null hypothesis
Laser *vs.* Wheel Lathe	0.4235	0.5154	Do not reject null hypothesis
Manual *vs.* Wheel Lathe	7.395	0.0067	Reject null hypothesis

Hence, for the flange slope parameter, there seems to be differences only in the spread of manual and wheel lathe measurements, with both pairs (laser *vs.* manual) and (laser *vs.* wheel lathe) not showing statistically significant differences in their variances.

3.5 Conclusions and Further Research

This chapter presented a systematic approach to compare the variances in measurement errors of different types of measurement devices in the context of railway wheelsets inspection, and in terms of three main dependent variables (responses): the change in flange thickness ΔFt, the change in flange height ΔFh, and the change in the flange slope ΔqR. A Linear Mixed Model (LMM) approach was conducted and compared to a simple linear model approach. The goal was to assess the agreement of measurements as a matter of the dispersion around the mean values. Since true values for the measurements were not known, it was assumed that the LMM would accurately represent the association

between the response variable and kilometers since last turning (kst) after controlling for the variation within unit number and kst.

Assessing agreement using inspection data from a system in operation under different conditions than the commonly adopted ones for controlled studies, such as repeated measurements and balanced data, is very rarely done in the scientific literature and has its limitations. On top of that, when the true measurement values are unknown, a lot of methods are of impractical use. Assuming a linear mixed model can explain the association between the response and the fixed predictors, and controlling for the variation of different groups by adding random intercepts, allows for comparison between measurement types by performing residual analysis, which was the methodology adopted.

Some relevant results can be drawn from the analyzes performed. The first relevant result relates to the comparison of the linear model (LM) with the linear mixed model (LMM). In this case, the LM considered only an initial wear and a wear rate as a function of kilometers since last turning (kst). The LMM considered an overall mean and the fixed effect of kst and it also included the variables kst and unit number as random effects. The LMM had a better fit in terms of a significant drop in RMSE when compared to the LM in all parameters studied. Moreover, the estimated standard errors for the residuals had a significant drop from the LM to the LMM model in all parameters, showing that ignoring the differences among different groups can potentially lead to incorrectly interpreting errors coming from the group structure of the data as measurement errors coming from the inspection devices. The LMM was shown to be a good approach to split the between- and within-unit variation, making it possible to compare the spread in the distribution of measurements coming from different devices.

The next relevant result is related to the behavior of the residuals assuming the underlying models were correct. Intuitively, measurements made with the gauge device (manual) and using the wheel lathe were expected to display a higher variance than the ones using the laser device, since they are more prone to human errors due to a bad positioning

of the gauge device or a bad positioning of the wheelsets in the under-floor turning wheel lathe. Within the discussion manual *vs.* wheel lathe, for the specific case of this train operating company, the variance of the manual measurements was expected to be higher, as represented by a prior flatter curve in Figure 3.2. This prior curve was based in the fact that a much larger number of technicians perfomed the manual measurements, while in the case of the lathe only the more experienced ones were allowed to do the turning actions, which would, at least in theory, reduce the inspector error.

In fact, analysis of the residuals considering a Levene's statistical test indicated that, in general, there is evidence in the data supporting small differences in the variances for the measurement devices, although those differences, by visual inspection of the boxplots in Figure 3.10,Figure 3.11 and Figure 3.12, do not seem something a maintenance manager should be concerned about. These differences were only statistically significant for the flange height parameter and for the comparison in the pair manual and wheel lathe measurements for the flange slope parameter. The statistical test for the flange thickness parameter did not reveal significant differences in terms of variances. Surprisingly, manual measurements showed a quite good performance, particularly relative to the laser equipment. One reason for that might be the fact that the number of observations using the laser equipment in the study is much smaller than the one of the remaining devices, since the train operating company started using the laser only in 2017, while the remaining observations of the other devices started in 2001. Another reason lies in the policy of inspection supervision in this company. A 'double-check' type of policy was employed in the observations coming from manual inspections and, after a gauge device inspection, the maintenance supervisor would check all the observations and compare them with a table against reasonable reference values of the wheel parameters according to the kilometers of the wheel. If any technician measured a really abnormal value due to an improper use of the device, the supervisor would identify it and repeat himself the measurement. This situation clearly favored the manual inspection in terms of reducing potential human errors in the

measurements of this device, improving its performance. Of course, this policy may not apply for other train operating companies, and generalizing the performance of the manual device based on the findings of this paper may not be straightforward.

In this study, the reproducibility of the inspector and the variability inherent to the human work, which are relevant aspects, were considered in the model indirectly, through the variances of the measurement devices themselves. However, the analysis in chapter 3 shows that some of the mechanisms related to the inspection activities inside maintenance shops may leverage the impact of the device performance. In addition, the flange thickness parameter seems to be less prone to differences in precision when compared to the other parameters. Therefore, it may be beneficial for IMs to use the flange thickness as a more robust parameter for the actual wheel condition. As future research, a study that incorporates the effects of human reliability in the inspection activities is recommended.

In a nutshell, there is evidence on the data, based on statistical tests, of differences in the variances of the meaurements obtained by these three procedures. Quantifying the effect of different precisions obtained by these three procedures is crucial for more efficient maintenance in railway wheelsets. For example, chapter 2 assumed ”perfect” inspections, i.e., no uncertainty around the measurements of the wheels used as model inputs. If uncertainty is incorporated, e.g. in a framework of models such as hidden or partially observable Markov decision processes, the modeling of ”imperfect” inspections may allow for a better estimate of the economic savings of changing the precision in equipment used for decision-making in the medium/long-term.

CHAPTER 4

SPATIOTEMPORAL APPROACH FOR RAILWAY TRACK MAINTENANCE

4.1 Introduction

Railways play a pivotal role in transportation systems worldwide. Effective maintenance policies are crucial to guarantee the reliability and profitability of such systems. Existing standards, such as EN 13848 [1] and UIC 714 [2], provide guidelines for maintenance activities of railways, promoting interoperability and harmonization across different companies operating in different countries. In this sense, these standards are often conservative, as they do not take into consideration the particularities of each train operating company, such as environmental conditions, track geometry, and design (e.g.: materials, curvature, track slope, the existence of substructure shortcomings). According to Peng et al. [82], for large-scale complex problems, due to lack of systematic solution techniques, current track maintenance practice in the railway industry mostly relies on ad-hoc trials, intuition, and experience of experts. However, as practical as a manual plan made by an expert may sound, there are many reasons why the scheduling process should not be made manually. It is time-consuming [83], some important constraints might be neglected or overlooked, and it is also a process that is highly dependent on the experience and judgement of the maintenance experts, in the sense that the company must have reliable knowledge management and training program to make new professionals capable of planning without significant loss to the process [83]. The conclusions and solutions developed by different planners vary greatly even given the same set of facts and criteria [84].

From the financial perspective, track maintenance is relevant as it involves high capital costs, with billions of dollars spent every year on track maintenance [85, 83], and

maintenance cost corresponding up to 35% of the total cost of operating the network [86]. Furthermore, poorly planned maintenance is likely to increase the costs, as not only the travel costs will be higher, but also other costs may occur, for example, multiple simultaneously ongoing projects within adjoining subdivisions that may severely block train traffic and cause high shipment delays [83]. Finally, in the case of faults that result in speed reduction or traffic blockage, the effects on operations are profound, as they incur extra crew costs, reduced equipment availability, and potentially, cargo loss [84].

The track geometry is defined by its layout and by track irregularities [87]. Fault detection and diagnosis of track geometry is also a challenging problem since the track is subject to fluctuating load and stress conditions, all involving randomness. Most railway infrastructure managers (IMs) assess the track quality either by comparing isolated track defects with predefined limit values and/or by computing key performance indicators (e.g. standard deviations) of track irregularities [88]. Such point-wise comparisons may be flawed since they ignore whatever has happened to the track up to that point and they also ignore the fact that the railway track is a very dynamic system, where faults can arise as a combination of various factors taking place at the same time/location. For example, Li et al. [88] point out the importance of considering dynamic responses at the wheel-rail interface at different train speeds and loads. They claim that some defects, such as rail welds, dipped joints and hanging sleepers, are difficult to detect by evaluating only the measured track data and illustrate with an example where geometric irregularities of short wavelengths ($\lambda = 0.5 - 3$ m) generate high dynamic wheel-rail forces, although the irregularity amplitudes are below the limit values specified in standards.

A deeper understanding of the features associated with track deterioration and the impacts of maintenance in the reliability of the system plays a fundamental role in a better maintenance decision-making process, allowing the train operating company to move to the next step in terms of enhancing its competitiveness. This can be achieved by creating a maintenance framework that makes a more accurate characterization of the remaining

lifetime distribution of the assets and can successfully translate this into the whole system reliability considering its dynamic and correlated nature. This framework also aims to provide the IMs information and flexibility for defining and changing the maintenance window according to schedules that exploit potential economic savings of maintenance decisions for single track sections versus grouping track sections together.

4.2 Related Work

Scheduling of railway track maintenance is a challenging task. In the literature, the track maintenance scheduling problem (TMSP) was addressed in many different formats. Some commonly adopted methodologies include applications in the field of operations research, such as MDPs, simulation, use of heuristics methods combined with linear and/or mixed-integer programming, etc. Others are more statistical-based approaches to the modeling of reliability and failure time distribution, such as Cox Proportional-Hazards Model - CPHM, Life-cycle cost analysis - LCCA. Some intrinsic characteristics of track maintenance jobs such as geographic dispersion, high set up times (including dislocation times), occupancy of track segments causing temporary unavailability, train delays, uncertainty regarding the best ”work window” for maintenance, variability in climate and environmental conditions and safety concerns may partially explain the complexity of the problem and the difficulty to come up with an appropriate model.

Reliability must be determined in a specific and operational sense [89]. In other words, failures should be defined according to the function that the system is expected to perform. In terms of railway track, a failure could be considered from a train speed restriction to a total infeasibility of circulation, both due to poor track conditions. Inspections, manual and automatic measurements are essential to gather knowledge about track, as it deteriorates over time. According to Berggren [90], the main processes of track deterioration are wear, fatigue and settlement. Periodic inspections are widely used as an instrument to identify possible defects and enable a better forecast of the reliability of the assets [13].

Some defects are visible, such as rail defects and track geometry irregularities, but the most important inspections are ultrasonic testing for the rail and track geometry measurements for the track [90], although some on-line continuous monitoring methods such as acoustic emission (AE) have recently grown in the number of applications [91]. Therefore, condition-based maintenance is very important for the track, as it indicates the necessity of intervention and the priority at which that the work must be carried on in order to avoid unacceptable conditions, such as speed restrictions or line closure [17].

From the economic point of view, the choice of maintenance interval aims to balance the trade-off between maintenance costs and costs originated from having a low-quality track, including, but not limited to: availability, the probability of incurring costs of an unplanned failure and the costs of an accelerated deterioration rate of the rolling stock that passes by the track. Railways operate under the conflicting objectives of minimizing infrastructure expenses while continuing to provide adequate service [86]. Therefore, the idea is to perform maintenance before the track deteriorates into an unacceptable condition, but not too early to the point of not being cost-effective, i.e., incurring costs of losing useful life. Thus, economic and safety effects of failures should be appropriately evaluated. For instance, since the replacement of most track components may cause the track to be unavailable until the work is finished, major maintenance (involving possibly all track subcomponents) might be preferred over a single replacement, cost and quality-wise. Besides, high logistics costs corroborate with the adoption of a plan-ahead strategy, which is commonly a year [92], unless a failed component is considered to be a safety hazard, then it should be replaced immediately [13].

Nonetheless, literature and field experience also shows that, even when a repair is performed, it is subject to the inherent randomness of stochastic processes, as well as degradation rates and equipment aging. For example, Quiroga and Schnieder [92] report an extremely fast degradation of track after tamping maintenance took place in a French railway. The authors cite some reasons why that might have happened, including water

under ballast, adverse weather conditions, or poor intervention quality. Such observations raise multiple questions to be taken into consideration before a maintenance strategy is adopted, for example, which level of intervention will bring the most cost-effective solution? How long is it worth to wait for major maintenance to happen before the chances of a failure are not negligible anymore? Similarly, if it was possible to estimate how much a given maintenance would add in terms of track's useful life, what would be the best timing strategy to reduce costs while keeping a given reliability goal?

Many works have attempted to study each of the track subcomponents and propose individual maintenance strategies for them: [13] focus on sleepers, [93] and [92] deal with tamping machine scheduling for the ballasted track, [94, 6, 95] also address the study of maintenance limits for tamping intervention, [96] analyze the degradation of turnouts.

The majority of applications in the field of operations research makes use of heuristics, as directly solving large-scale time-space network problems with many side constraints is well known to be computationally intractable [82]. When dealing with a highly complex model, for which no standard optimization method can be directly applied, there are two possibilities: either adapt the model (usually by simplifying it) to enable the use of a standard optimization method or come up with a new solving method to fit the model [92].

A classic work by Higgins [97] uses Tabu Search (TS) for the short-term scheduling of track maintenance activities at Queensland Rail (Australia), where the objective function is to minimize the combination of expected interference delays with the train schedules and prioritized finishing times, using cost budget as a hard constraint. Oyama and Miwa [93] develop a model for measuring the degradation of track surface and predicting the effects of future maintenance (restoration) for then applying all-integer linear programming (AILP) to determine the optimal maintenance schedule for the Multiple Tie Tamper machine (MTT), so that the total improvement of track irregularities (measured by the sum of restorations) is maximized in a period of one year. The model specifies, for the optimal operation of the tamping machine, several parameters, including: which depot the MTT

should be located, where the restoration must be executed, the timing, etc. Lake et al. [86] develop a model for short-term scheduling of maintenance activities based on the simulated annealing metaheuristics, taking into consideration individual set-up and take-down times and restrictions on the times maintenance activities are permitted (due to crew work hours). They also consider the possibility of discontinued maintenance. Zhao et al. [13] use integer programming combined with the steepest gradient method to minimize the number of sleepers restored during maintenance work, subject to meeting the minimal requirements of safety and reliability.

Peng et al. [82] present a time-space network model to solve the TMSP through an iterative heuristic solution applied to a large-scale problem at CSX Railway (USA), with the objective being to minimize the total travel costs of maintenance teams. Andrade and Teixeira [98] present a bi-objective integer optimization model based on the Simulated Annealing (SA) metaheuristic for planning maintenance and renewal actions related to track geometry in a railway network, balancing costs with train delays. Vale et al. [99] propose a mixed-integer programming model to optimize the maintenance schedule for tamping of railway tracks.

Other works use optimization to address the maintenance team scheduling problem, for example [15, 83].

Mixed-approaches encompassing statistical degradation modeling and further optimization of the maintenance policy are also presented in the literature. Quiroga and Schnieder [92] develop a heuristic-based algorithm for the tamping machine scheduling, validated by a real application in a French high-speed line. The authors develop a track geometry deterioration forecasting method that relies on two assumptions: first, the degradation value achieved after the n^{th} tamping intervention is normally distributed and, second, the evolution of the degradation value between the two tamping activities can be described by an exponential function. Mean and variance are then estimated. Next, a heuristic scheduling algorithm is solved to find a feasible solution consisting of a set of N

interventions (one per night) which minimizes the expected longitudinal leveling at the next campaign. The authors also emphasize the importance of wisely choosing the objective function as it should express the objective of the railway track maintenance process, which is very particular and varies from one company to another. Some examples of objective functions are the total reduction track geometry deviation, the expected time to failure and the implemented expected longitudinal leveling at the next campaign.

Zhang et al. [17] consider the uncertainties of the deterioration process, the safety of transportation service, the lifetime loss of the replaced track, the maintenance cost and the travel cost, to develop a heuristic approach with the genetic algorithm (GA) for searching for a maintenance schedule that minimizes total costs over a finite planning horizon. They use the track geometry data obtained from the measurement train to establish the condition of each track segment and to derive a probability distribution. The deterioration is then classified into three categories: good condition, degraded condition awaiting maintenance and an unacceptable condition, where the increased circulation costs of the last one make it a maintenance priority.

Shafiee et al. [100] present an optimal bivariate (age-usage) maintenance strategy for railway tracks that allowed the average long-run maintenance costs per unit time to be minimized. Statistical modeling of age and usage is performed.

These studies reinforce the power of optimization techniques in solving related track scheduling problems and delivering fast and reliable solutions, especially when combined with statistical methods for degradation modeling. On the other side, many ”purely” statistical-based approaches have succeeded in exploring some inherent features of the track correlated nature, often taking advantage of condition-monitoring data. Some of them will be summarized below.

Arasteh-khouy et al. [94] describe the tamping maintenance strategy of Trafikverket (Swedish Transport Administration). Data from an inspection car is used to monitor track quality and some indices based on the standard deviation of vertical and lateral

displacements are provided. Tamping (which compacts ballast and corrects track geometry faults) is performed according to condition-based maintenance and comparison to the alert limits: results above the maximum allowed deviation from design condition will require corrective maintenance, and preventive maintenance is performed within narrower limits. Statistical analysis revealed a high frequency of failures in certain locations, independently of how much maintenance was executed, indicating that tamping was not effective to remove the root cause of failure in some cases. The authors claimed that cutting the capital cost required for fixing the root cause of failures can result in large maintenance costs for years afterward to compensate for the track substructure shortcomings. Other interesting results were related to different geometry parameters (such as twist and longitudinal level) showing different trends over time, seasonal climate and temperature having an effect on failure rate and failure rate among track segments not being uniform.

In another work, Arasteh-khouy et al. [95] discuss the track geometry inspection interval. A Swedish railway iron ore line subject to extreme weather conditions and high axle loads was studied. The model assumed that both corrective and preventive tamping were performed based on the results of fixed-time inspections and the objective was to identify the interval and frequency that would minimize the total cost per unit of traffic load (MGT) for any length of the track section, i.e., inspections should be performed only when their costs were offset by a resulting reduction in expected future costs. Probability distributions faults were estimated, with segments without any fault occurrence taken as right-censored data, and linear regression was used to rank different probability distributions. The results pointed out that expanding the inspection interval from 2 to 4 months would decrease total maintenance costs, which was partially explained by the fact that the majority of track segments had shown slow degradation rates. Finally, the study made a recommendation to increase the number of inspections in some track sections with higher degradation rates, in order to reduce risk and ensure the safety level.

A Bayesian model framework to assess and update the railway track geometry

degradation uncertainty throughout its life-cycle is proposed by Andrade and Teixeira [101]. They show that the uncertainty associated with deterioration rates for track sections at their design stage is very high, but it reduces significantly as operation starts and more inspection data is collected so that it is recommended a 2-3 years "warm-up" period for better maintenance assessment and planning. In two other works [102, 103], the authors expand some of these ideas by exploring hierarchical Bayesian models to predict track geometry degradation, paying particular attention to the need of inserting spatial statistical dependencies between model parameters to tackle spatial interactions between consecutive rail track sections. Lastly, in related work, Andrade and Teixeira [104] explore a quantitative model to assess the effect of different alert limits for quality indicators in track maintenance. They propose a simplified deterministic Markov policy (similar to the inventory model problem with two decision rules) that is able to compute optimal alert limits for different track speed sections.

Li et al. [88] claim that current assessment methods of track geometry that do not give enough attention to the wheel-rail interface may not be sufficient for the establishment of an effective maintenance policy or even train speed settings. For the authors, wavelength contents should be added to the analysis of dynamic wheel-rail forces, in order to develop a model that can successfully identify track sections that are likely to produce high track forces and unsafe vehicle responses. In this study, a track section of a Swedish railway is monitored by a vehicle with only wheelset, car body, and primary suspension, so that data from wheel-rail forces caused only by track irregularities can be obtained. Numerical simulations were carried out and when track irregularities are filtered so that only short wavelengths ($\lambda = 0.5 - 3m$) remain, a more linear relationship between force and vertical track irregularity can be observed, showing that the method is effective to assist the monitoring of track condition and, therefore, allowing for better maintenance planning.

As shown above, both optimization and statistical-based approaches have pros and cons. Optimization models often involve several simplifying assumptions for the purpose

of computational tractability, which may not be realistic in practice, resulting in many models simply ignoring discrepancies found on the data. Moreover, accounting for the many possible decision variables and constraints involved in the railway track maintenance planning is a highly non-trivial task. At the same time, understanding the degradation process is mandatory if the company wants to predict (or have a good estimation of) the correct time for inspection, maintenance and renewal [6]. The aforementioned data-driven approaches put into evidence the importance of routine inspection for track maintenance planning. Visual inspection, automated methods to detect variations in geometry, flaws in rail are commonly used to gather data regarding track condition. They support not only short term but also long term maintenance plans, especially because an analysis of past inspections can determine the growth trend of the various track deficiencies and lead to the planning of a major rail replacement program [84].

Track deterioration is a stochastic process, which is affected by a variety of interrelated factors [105]. Hence, completely ignoring the track features, as well as the serially correlated nature of track sections, may lead to simplistic models. The methodology explored in this chapter combines the stochastic modeling of track irregularity degradation with an optimization tool (in the form of an MDP model) to provide a framework for modeling track faults and exploiting potential economic savings of maintenance decisions for single track sections versus grouping track sections together.

The methodology consists of fitting a Kriging model based on a Gaussian correlation function which captures the correlated nature of the consecutive track sections. Only a few studies have explored the use of the Kriging approach in the context of railways, and most of them do not deal with degradation modeling directly. A couple of exceptions are the works by Cremona et al. [30] and Bergquist and Söderholm [106]. In particular, Cremona et al. [30] deal with the problem of railway wear prediction and uses a Universal Kriging model to exploit the dependence of wear coefficients with similar contact pressure and sliding speed, creating a continuous wear coefficient map which is more informative

than the currently available in the literature. In terms of track maintenance, Bergquist and Söderholm [106] use condition-monitoring data fitted into a Kriging model to evaluate the possibility for earlier detection of track faults using both temporal and spatial information.

Nonetheless, the railway track is a distributed system that is affected by different heterogeneous factors, with uncertainty being the major characteristic of its behavior [8]. Some of these factors cannot be measured directly by the monitoring variables or are simply unknown, although they still impact the track quality. For example, Shenton [107] demonstrated that track condition is strongly dependent on the initial level (design) after implementation and maintenance activity cannot change it. Öberg and Andersson [108] showed that different vehicles and loads lead to different degradation behaviors. Dahlberg [109] concluded that load smaller than a specific threshold can not cause track settlement, whereas slightly increased loads cause linear settlement and excessive loads cause settlement to be non-linear. Andrade and Teixeira [101] showed that uncertainty is significantly high at the design state of the track, when compared to other stages (e.g. after the first inspection, between the first inspection and the first tamping, etc.). Quiroga and Schneider [110] observed that the longitudinal level showed accelerated growth in initial tamping, but this behavior decreased with more tamping interventions.

The statistical modeling approach used in this chapter is a Bayesian methodology to analyze correlated data, often measured with error. Metamodeling with Kriging is useful as it serves as a core building block for interpolators, even when only a minimal amount might be known about the output function [111].

4.2.1 Main Contributions

This chapter focuses on track geometry faults. Spatial-time data from a railway company operating in Brazil is considered. A transformation of the twist standard deviation, which is a commonly used statistic to measure track geometry condition, is calculated based on data and a spatiotemporal Kriging approach with a Gaussian correlation function is proposed

by taking advantage of regular grid data. The first main contribution of this work is the utilization of the Kriging technique to model the spatiotemporal evolution of a railway track safety indicator, with further comparison of two Kriging models: Ordinary (OK) and Limit Kriging (LK), both built with a Gaussian Correlation function. The underlying statistical approach is conceptually simple to implement, as the degradation process function does not need to be specified.

The second main contribution is the integration of the statistical approach with a control chart approach (CUSUM) and an optimization solver (MDP) is proposed, so that a more comprehensive decision framework for track maintenance planning is provided. In light of this, the third main contribution relates to applying the integrated decision framework through an MDP approach to assess the cost percentual differences arising from different grouping strategies of track section, while also exploiting potential economic savings of maintenance decisions for single track sections versus grouping track sections together.

4.3 Preliminares

4.3.1 Kriging

To statistically model the spatiotemporal characteristics of track geometry degradation, an interpolation technique denominated Kriging is used. It was originated in geostatistics but it is now widely used in many fields, including spatial statistics [112].

Kriging performs an interpolation based on regression against observed data points, weighted according to a data-driven covariance function, which is what differentiates it from other interpolation techniques such as splines, polynomial interpolation, radial basis interpolation, inverse distance weighting, etc. While Kriging may be more complex than these other techniques, it takes advantage of covariance between samples and, therefore, has greater potential to generate better models, since the nature of data, sampled along a railway track, makes measurements next to each other autocorrelated [113]. It is almost intuitive to imagine, for example, that if the track is degraded at some point, it

is very likely that the surrounding points will also have some level of degradation. As emphasized by Uzarski and McNeil [84]: "individual distresses may increase in size and may multiply when component elements (for example, individual ties) fail and loads redistribute themselves throughout the structure and overload other component elements". Thus, a failed portion of the track translates into the surrounding portions likely suffering overload to compensate for the failed one. It is also intuitive to think that the correlation exists in time domain as well: if a portion is degraded, unless it is under influence of extreme stochastics events (e.g. accidents) or abnormal usage, it is going to be as or more degraded some time afterwards, if no repair is performed [113].

Under spatial autocorrelation, closer points are more likely to display similar values, which usually allows the interpolation to give reliable estimates at nearby locations in the absence of observed values. If one ignores autocorrelation and decides to use conventional statistical methods that rely on independence assumptions, such as linear regression, the results and conclusions may be erroneous.

Accounting for the uncertainty concerning wear and fault prediction is also advisable. In this sense, some metamodels built with Gaussian Process and Kriging can be useful as to provide a good approximation of the behaviour of a complex system and to study how different the impact of the different settings of inputs in the predicted output values.

For the above reasons, Kriging was chosen to model track degradation data, exploiting the similarities among different track segments and different inspections. Following Joseph [112], the basic spatial Kriging model can be state according to Equation 4.1:

$$Y(\mathbf{x}) = \mu + Z(\mathbf{x}) \tag{4.1}$$

where $\mathbf{x} \in \mathbb{R}^p$, μ models the large scale variation and $Z(\mathbf{x})$ is assumed a weak stationary stochastic process with mean 0 and covariance function $\sigma^2 \boldsymbol{R}$, see [112].

(Weak) Stationarity is needed to allow for inference. Take $Y(\cdot)$ as a Random Function (stochastic process) and consider $y(\cdot)$ as one draw (realization) of this RF, say $y(x) =$

$Y(x, w), w \in \omega$, where ω is the sample space of elementary outcomes and $Y(\cdot, w)$ refers to a particular function from X to $\mathbb{R}^1$ (see [111], chapter 2). In classical statistics inference about certain population would be made by taking multiple draws of $Y(\cdot)$ with multiple w. However, for the Kriging application, training data $y(\mathbf{x})$ are the values $y(\mathbf{x}) = Y(\mathbf{x}, w)$ corresponding to a single w and, hence, stationarity is required to make inference valid [111]. Because of its analytical tractability, Gaussian Process (GP) model is usually preferred for generating function draws [111].

It is well known that any Gaussian process is completely determined by its mean function and covariance function. Therefore, a GP $Y(\cdot)$ is said to be stationary if for any $n \geq 1$, any $\delta \in \mathbb{R}^p$, any $x_1, \ldots, x_n \in$ (some subset) $X \subset \mathbb{R}^p$ and any $x_1 + \delta, \ldots, x_n + \delta \in X$, $(Y(x_1), \ldots, Y(x_n))$ and $(Y(x_1 + \delta), \ldots, Y(x_n + \delta))$ have same mean vector and same covariance matrix, with covariance depending only on the distance between inputs, e.g., $\text{cov}(Y(x_1), Y(x_2)) = c(x_1 - x_2)$, where $c(\cdot)$ represents the covariance function. Thus, for any arbitrary point it holds that $\text{cov}(Y(x), Y(x)) = c(x - x) = c(0) = \sigma^2$ (constant variance, homocedasticy) and $\text{correl}(Y(x), Y(x)) = \frac{c(x-x)}{\sigma\sigma} = \frac{c(0)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = R(0) = 1$. Lastly, it must also hold that covariance and correlation are symmetric about the origin, i.e., $c(x) = c(-x)$ and $R(x) = R(-x)$.

Returning to the Kriging model in Equation 4.1, two variations of this equation will be considered. In the first one, known as Ordinary Kriging, μ is constant but must be estimated from data. In the second variation, Limit Kriging [112] μ is a function of x, which partially overcomes one of the aspects of Ordinary Kriging of having predictions going towards μ, allowing for more accentuated "peaks" of the response variable.

Given data $(x_i, y_i), i = 1, \ldots, n$, the equation for the Ordinary Kriging predictor is given by Equation 4.2:

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \boldsymbol{r}(\mathbf{x})' \boldsymbol{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}) \tag{4.2}$$

with the estimate for μ using Generalized Least Squares being:

$$\hat{\mu} = \frac{\mathbf{1'}\boldsymbol{R}^{-1}\mathbf{y}}{\mathbf{1'}\boldsymbol{R}^{-1}\mathbf{1}} \tag{4.3}$$

Ordinary Kriging predictor is also known to be the Best Linear Unbiased Predictor (BLUP), which minimizes the mean squared prediction error $\mathrm{E}\left[\hat{Y}(\mathbf{x}) - Y(\mathbf{x})\right]^2$ under Equation 4.1.

Next, the equation for Limit Kriging can be obtained as:

$$\hat{y}(\mathbf{x}) = \frac{\boldsymbol{r}(\mathbf{x})'\boldsymbol{R}^{-1}\mathbf{y}}{\boldsymbol{r}(\mathbf{x})'\boldsymbol{R}^{-1}\mathbf{1}} \tag{4.4}$$

where $\boldsymbol{r}(\mathbf{x})'$ is the transpose of a $n \times 1$ vector with the i^{th} element being $\boldsymbol{R}(x - x_i)$, $\boldsymbol{R}$ is the $n \times n$ correlation matrix with ij^{th} element $\boldsymbol{R}(x_i - x_j)$, $\mathbf{1}$ is the $n \times 1$ vector of ones.

A simple proof involving Cauchy-Scharwz Inequality [112] shows that the Mean Squared Prediction Error for the Limit Kriging Predictor is greater than the one for Ordinary Kriging (which is also the BLUP) with constant mean. However, Limit Kriging predictor can potentially outperform Ordinary Kriging one as it is more robust against mispecification of parameters, i.e., if the model in Equation 4.1 is not reasonable, Ordinary Kriging predictor will ”pull” predictions towards the mean, whereas Limit Kriging is able to overcome this problem [112]. This feature may be particularly suitable here as it is desirable to find out-of-control points for the degradation data, which indicate possible future failures of track and, therefore, it is interesting to compare a more conservative (Ordinary Kriging) with a more ”radical” approach (Limit Kriging).

As showed above, Kriging exploits covariance as a distance function. The correlation function $R(\cdot)$ is chosen to be Positive Definite (PD), which will allow for the covariance matrix to be nonsingular. The choice of Gaussian correlation function is, therefore,

reasonable and given by:

$$\boldsymbol{R}(x_i, x_j) = \exp\{-\sum_{l=1}^{p} \theta_l (x_{il} - x_{jl})^2\} \tag{4.5}$$

where $\theta \geq 0$ are the unknown correlation parameters that must be estimated. For the present work, these estimates will be obtained by maximizing the likelihood, or minimizing the loglikelihood, as shown in Equation 4.6.

$$\min_{\theta} n \log \hat{\sigma}^2 + \log |\boldsymbol{R}| \tag{4.6}$$

where $\hat{\sigma}^2 = (\mathbf{y} - \hat{\mu}\mathbf{1})' \boldsymbol{R}^{-1} (\mathbf{y} - \hat{\mu}\mathbf{1})/n$.

The Gaussian correlation function produces a local trend that is infinitely differentiable, a smooth curve to deal with spatial autocorrelation (for more details, see [114]).

In this study, however, the goal is also to be able to predict in the time domain, hence, temporal autocorrelation is also crucial. The spatial Kriging model above can be modified to account for time as well. As mentioned earlier, one can take advantage of spatiotemporal data falling onto a regular grid (i.e. separable data). By using a spatiotemporal approach, the computer tractability of the problem can be improved substantially. This is important since the estimation of the correlation parameters (in Equation 4.6) involves the inversion of the correlation matrix, which can be very large, possibly leading the estimation to be numerically unstable [115]. For observations collected in a regular grid, this problem can be overcome by applying Kronecker product formulation for constructing the correlation matrices (i.e. $\boldsymbol{R} = \boldsymbol{R}_x \otimes \boldsymbol{R}_t$, where $\otimes$ stands for the Kronecker Product, x represents space and t represents time), which also converts the spatial model into a spatiotemporal approach. In order to apply it, the above equations are going to be modified the following way (for more details, see [115]).

The Kriging model in Equation 4.1 becomes in the spatiotemporal approach:

$$Y(\mathbf{x}, \mathbf{t}) = \mu + Z(\mathbf{x}, \mathbf{t}) \tag{4.7}$$

Then, the equations for the Ordinary Equation 4.2 and Limit Equation 4.4 Kriging predictors can be updated, respectively, by :

$$\hat{y}(\mathbf{x}, \mathbf{t}) = \hat{\mu} + (\boldsymbol{r}_x(\mathbf{x}) \otimes \boldsymbol{r}_t(\mathbf{t}))'(\boldsymbol{R}_x^{-1} \otimes \boldsymbol{R}_t^{-1})(\mathbf{y} - \hat{\mu}\mathbf{1}) \tag{4.8}$$

$$\hat{y}(\mathbf{x}, \mathbf{t}) = \frac{(\boldsymbol{r}_x(\mathbf{x}) \otimes \boldsymbol{r}_t(\mathbf{t}))'(\boldsymbol{R}_x^{-1} \otimes \boldsymbol{R}_t^{-1})\mathbf{y}}{(\boldsymbol{r}_x(\mathbf{x}) \otimes \boldsymbol{r}_t(\mathbf{t}))'(\boldsymbol{R}_x^{-1} \otimes \boldsymbol{R}_t^{-1})\mathbf{1}} \tag{4.9}$$

with mean and variance redefined analogously. Lastly, the likelihood in Equation 4.6 shall be modified to :

$$\min_{\theta} mn \log \hat{\sigma}^2 + n \log |\boldsymbol{R}_x| + m \log |\boldsymbol{R}_t|, \tag{4.10}$$

since $|\boldsymbol{R}| = |\boldsymbol{R}_x| \otimes |\boldsymbol{R}_t| = |\boldsymbol{R}_x|^n |\boldsymbol{R}_t|^m$, n being the run size (in the spatial domain) and m being the number of observations from time domain in each run.

4.4 Spatiotemporal Approach Application to a Single Heavy-Haul Track Portion

4.4.1 Track Twist

In this chapter, information about track superelevation (or elevation of the rails) is used to compute two statistics for track geometry condition assessment: the cant (cross-level) and the twist. As shown in Figure 1.2, cant is the difference in elevation between the outer rail and the inner rail. It naturally occurs in curves, since it is necessary that the outer rail is slightly more elevated than the inner rail so that the train can make the curve. In a straight line, the cant is expected to be approximately equal to zero. For both curves or straight lines, however, the rate of change of cant should not be high, i.e. the rate of change of track

superelevation should not be abrupt, which could increase the risk of derailment [113]. This rate is called twist: the amount in height by which cant is increased or decreased in a given length of track. Twist is a type of track quality indicator commonly used among train operating companies and it can be monitored throughout the life of the track. In addition, many studies (a detailed review can be found in [8]) suggest the monitoring of the standard deviation of track defects as indicators of track's geometry quality. A transformation of the track twist standard deviation is proposed next as the main parameter for the spatiotemporal approach.

Twice a year a measurement vehicle that inpects the track collects data on every segment. The data includes the track twist measured in a 20 m track length (i.e. cant measured at 2 points 20 m apart from each other) for every foot of track. Measurements were collected approximately six months apart from each other, in a total of nine inspections (or a range of 4.5 years). Figure 4.1 shows the data collected in the first eight inspections, in ascending order according to the date of the inspection, where the y-axis corresponds to the twist (in mm/m) and the x-axis corresponds to the location of the measurement (in km).

The raw data, as shown in Figure 4.1, is further modified by the computation of the standard deviations for the twist measurements for every 200 m of track, in a total of 49 segments (from km 65.2 to km 74.8). Then, a transformation of the standard deviation of the track twist as suggested in [113] is applied, by first taking the difference in range of the twist standard deviations for each track segment. At this point, all data are positive. Next, a double natural logarithm transformation, i.e. log(log(range(sd(twist)))), is applied to the data, and the result is taken as the monitoring statistics. Shapiro-Wilk test reveals that normality assumption cannot be rejected at a significance level of 5%, as shown in Table 4.1. Normality here is not necessary for Kriging application, although it is necessary for constructing the confidence intervals that will be shown herein.

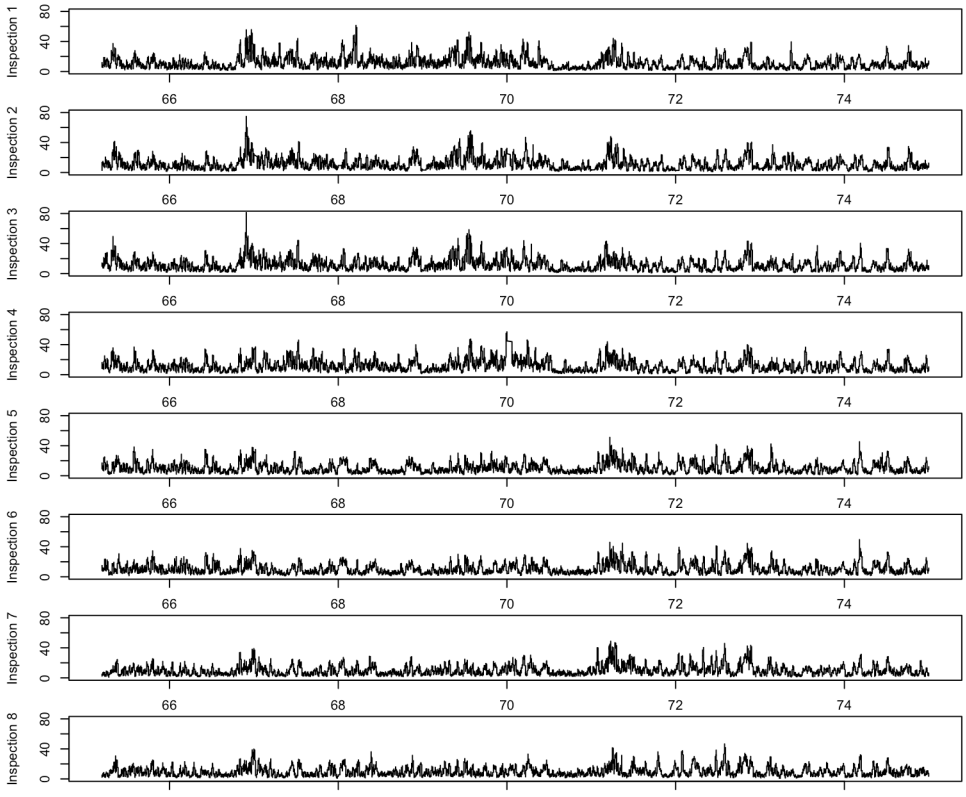

Figure 4.1: Twist measurements (in mm/m) per inspection

Table 4.1: Shapiro-Wilk test for normality

Test Statistic	p-value
0.9984	0.9613

4.4.2 Ordinary and Limit Kriging

As mentioned earlier, the nature of collected data makes it spatial and time autocorrelated. Intuitively, if track is degraded at some point in time and no event (such as maintenance activity) dramatically changes this situation, it is likely that the next measurement will point out a degraded track. Under these conditions, Kriging Models (Ordinary and Limit) were trained with 4-year data (a total of eight inspections) and applied to find one-step-ahead extrapolation in the time domain.

One parenthesis here is that Kriging is widely known by its power as an interpolator. Hence, the extrapolation here is more of an experimental investigation than an established

approach. Twist values tend to revert to a mean [113], as a negative twist is usually followed by a positive twist when the rail rises back after the deformed segment. Hence, the choice of Ordinary Kriging seemed appropriate. An alternative method with Limit Kriging was also proposed to check its performance, especially in the "peaks", or points where response gets out-of-control indicating possible future failure points.

Empirical Bayes estimation with a Gaussian correlation function was applied to get maximum likelihood estimates for the θ parameters for space and time (Equation 4.10), namely $\hat{\theta}_x$ and $\hat{\theta}_t$, and the estimates for mean and variance being: $\hat{\mu} = 0.385$, $\hat{\sigma}^2 = 0.0076$. Figure 4.2 shows the predictions (in blue) with Ordinary (a) and Limit (b) Kriging *vs.* real data (in green) and a two-sided confidence interval based on the Mean Squared Prediction Error (the red dashed line) for the Kriging predictions. Root Mean Squared Error (RMSE) was computed for both models, resulting in approximately 0.058 for Ordinary Kriging (OK) and 0.065 for Limit Kriging (LK). Although OK's RMSE was slightly lower than LK's, visual inspection of Figure 4.2 shows that LK has better performance in the possible failure points, because of its ability to get closer to possible extreme values of the twist, which is the main interest of this study. Hence, LK was chosen as final model.

The predictions above, although close to real values, seem not to be accurate enough in some specific points. One obvious reason is that possible maintenance activities were not considered, and hence the model does not incorporate repairs (e.g. tamping operations) or renewals to the track. Maintenance actions that affect twist include a long list of jobs, such as tamping, sleeper changes, ballast maintenance, alignments, among others, making the task of estimating each of these jobs' effect on the response very complicated. To minimize this obvious model's drawback, one of the main criteria for the selection of the portion of the track under study was that it had not undergone major maintenance works or renewals in the period studied, except for possible tamping and sleeper changes on some isolated portions, which could potentially explain why the Kriging model fails in some specific points. Moreover, another obvious difficulty related to modeling the repairs is the effect of

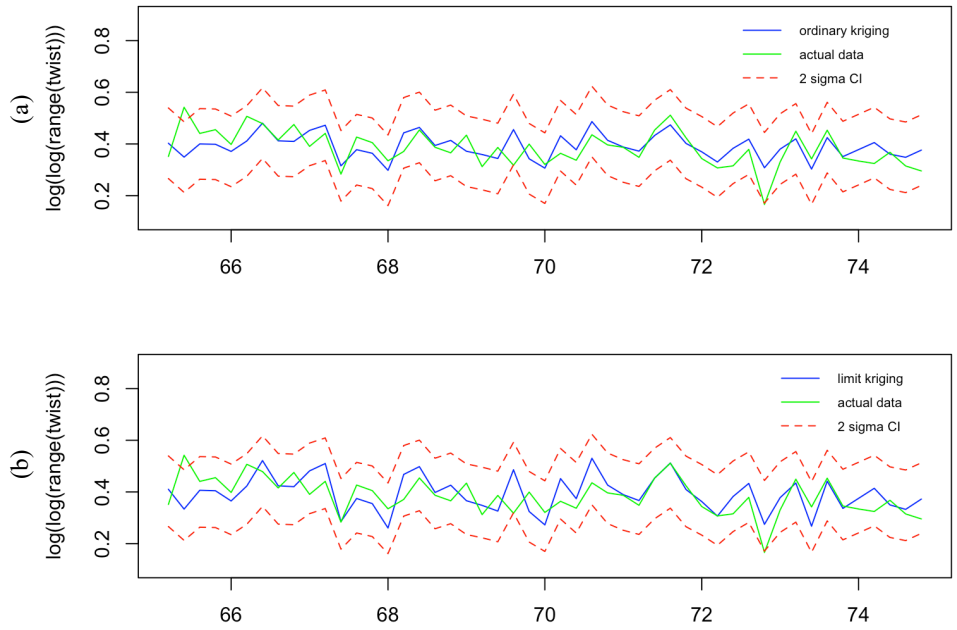

Figure 4.2: (a) Ordinary Kriging Predictions *vs.* real data and (b) Limit Kriging predictions *vs.* real data

tamping effectiveness, which decreases as more tamping interventions are performed, until tamping is not effective anymore and a renewal of the track is recommended to bring it back (closer) to its design state [92]. In this case study, information on past number of renewals and tamping activities was not available. Of course, modeling repairs and renewal activities can definitely enhance the accuracy of predictions, and this is left as future research.

Having the LK predictions, the next question to be answered is: which track sections should the maintenance plan focus on? To answer this, it is first necessary to define which track sections deserve attention which, in this case study, will be called out-of-control sections. Ideally, the range of twist in a track section should be close to 0, i.e. the section did not suffer abrupt changes in superelevation. However, the difference in the range of the standardized twist considering all track sections has the minimum of 3.13 (twist standard deviations) and maximum of 6.78 (twist standard deviations). In the double logarithmic transformation, this is equivalent to 0.13 and 0.64, respectively. Obviously, for this railway track, it is not realistic to consider that all segments should have standardized

range difference close to 0. Hence, the problem is: considering the current track situation, what are the points that could be considered out-of-control?

4.4.3 Statistical Process Control with CUSUM Chart

In terms of statistical process control (SPC), many techniques exist with the objective of detecting small shifts in the mean (and/or in the variance) values of some control variable. Charting techniques such as Shewhart Control Chart, EWMA and CUSUM have been widely used in the literature. However, many of these charting techniques are built under the assumption that the observations are independent from each other. Therefore, applying any of the above techniques in the Kriging predictions could result in an erroneous output, as there is presence of autocorrelation in the data under study. To overcome this problem, a similar methodology to the one proposed in [116] was used, enabling the application of the Cumulative Sum (CUSUM) Control Chart to the autocorrelated observations. In [116], observations were assumed to follow a first order autoregressive (AR(1)) model with an additional random error corresponding to measurement error. Letting X_{ti} represent the i^{th} observation from the process at time t, it is turns out that: $X_{ti} = \mu_t + \varepsilon_{ti}$, where μ_t is a random variable representing the mean of the process at time t, varying over time according to an AR(1) process. The random error ε_{ti} is assumed normal with mean 0 and variance σ_{ε}^2. The objective is detecting a change in the overall mean $\xi = \mathrm{E}[\mu_t]$. Since it is assumed that μ_t follows an AR(1) process then Equation 4.11 holds:

$$\mu_t = \xi(1 - \phi) + \phi\mu_{t-1} + \alpha_t \tag{4.11}$$

where ϕ is the correlation between μ_t and μ_{t-1} and α_t is a random shock which is normally distributed with mean 0 and variance σ_{α}^2. It is also assumed that $|\phi| < 1$ so that the process is stationary. Hence, the method proposes a "correction" for the observed values, taking into consideration the autocorrelation between observations. Considering the X term that forms control statistics for CUSUM, it would be transformed to: $\bar{X}_t = \mu_t + \bar{\varepsilon}_t$.

The methodology described above has a lot of similarities with the Kriging model utilized earlier. Hence, instead of an AR(1) process, the above expressions will be modified considering the Kriging model in Equation 4.7, where errors arise from spatial and time autocorrelation (using the Gaussian Correlation function in Equation 4.5). The X term that forms control statistics for CUSUM would be simply the Kriging model predictions at each time, for each track section (still assuming an overall mean μ).

Since the goal is predicting in time domain, CUSUM is applied to each track section (49 in total), each with nine values over time, corresponding to the nine inspections/predictions. Moreover, the interest is to detect increase in the mean, as an increase in the difference of the range of the twist is related to a more degraded track. Therefore, a one-sided CUSUM chart is appropriate. The control statistic for detecting an increase in the overall mean μ is:

$$Y_{ts} = \max\{0, Y_{t-1,s} + \bar{X}_{ts} - \mu_0 - K\} \tag{4.12}$$

where s represents track section $s = 1, \cdots, 49$, t is the time $t = 1, \cdots, 9$, $\bar{X}_{ts}$ is the Kriging prediction at section s and time t, μ_0 is the in control mean which was taken to be 0.38 and K is the allowance or slack value, which was taken to be 0.07 (or about half of the difference of what would be an out-of-control process mean and an in-control process mean). The decision rule for CUSUM control chart is: whenever the control statistic Y_{ts} exceeds the decision interval H, the process is considered out-of-control.

In this application, four different values $(3, 3.5, 4, 4.5)$ for the decision interval H were tested, based on standard deviations for twist. For example, if a maximum of 3 standard deviations from the overall mean is allowed, then $H = \log(\log(3)) = 0.094$, hence any control statistics exceeding this will be considered out-of-control. The other values for H can be computed analogously. For example, for the aforementioned H, Figure 4.3 shows the out-of-control track sections (dots) on time period 9, using the limit Kriging model.

As expected, increasing H results in less track sections being considered out-of-control. The H-to-pick will depend on the railway Infrastructure Manager (IM), who should pick a

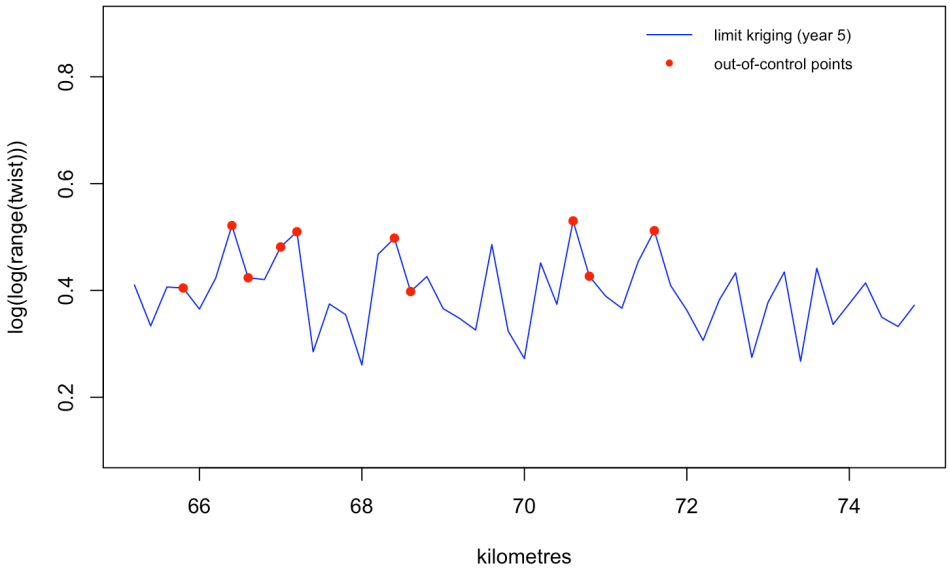

Figure 4.3: Out-of-control track sections for $H = 0.094$

threshold that, at the minimum, guarantees safety of operations.

Considering the four suggested values for H, and under some other assumptions, it is desired to know, under a limited budget, and considering long term strategy (10-11 years look ahead) in terms of maintenance plan, what would be a reasonable maintenance policy for the track sections under study. A Markov Decision Process is applied and optimal results in terms of minimal cost policy are discussed in the context of the MDP in the next subsection.

4.4.4 Markov Decision Process Approach

According to Ross [44], a stochastic process $\{X_n, n = 0, 1, 2, \cdots\}$ that takes on finite number of possible values and for which the conditional distribution of any future state X_{n+1}, given the past states $X_0, X_1, \cdots, X_{n-1}$ and the present state X_n, is independent of the past states and depends only on the present state is called a Markov Chain. Whenever the process is in state i, it is assumed that there is a fixed probability p_{ij} that it will next be in state j. For further details on Markov Chains, refer to [44]. An MDP is a sequential decision process for which the decisions produce a sequence of Markov Chains

with rewards, as discussed in chapter 2.

Under each one of the maintenance policies, which here will vary according to the percentage of out-of-control points allowed in the given portion of track, a sequence of states $\{X_n, n = 0, 1, 2, \cdots\}$ will constitute a Markov Chain with transition probabilities p_{ij}. For every state i, an action a is chosen from the set of possible actions A (assumed finite), and for each action a set of state and action dependent costs $r(i, a)$ is established. Hence, if the process is in state i at time n and an action a is chosen, the next state of the system is determined according to $p_{ij}(a)$. By letting X_n denote the state of the process at time n and a_n the action chosen at time n, the above is equivalent to:

$$\Pr(X_{n+1} = j|X_0, a_0, X_1, a_1, \cdots, X_n = i, a_n = a) = \Pr(X_{n+1} = j|X_n = i, a_n = a) = p_{ij}(a) \tag{4.13}$$

Thus, the transition probabilities are functions only of the present state and the subsequent action, where the actions are defined according to the policy. A simple application of a MDP will be used to get expected total costs arising from each strategy, allowing for the railway specialist to know the best policy in terms of costs associated with each decision interval H.

For the MDP model, a total of 11 states are proposed, with the first state representing 0 out-of-control track sections, the second corresponding to 10% out-of-control track sections, the third corresponding to 20% out-of-control track sections, up to 100% out-of-control track sections (state 11). For each H, 9 different maintenance policies are studied: strategy 1 being repair as soon as only 20% sections are out-of-control, increasing by 10% for each strategy, with strategy 10 being repair only if all track sections (or 100%) are out-of-control. This means that the transition probabilities from state i to state 1 (design state) will be defined according to the maintenance policy, e.g. under strategy i, $p_{i1} = 1$. The set of possible actions considers the grouping strategy in terms of percentages of track sections

to be repaired, i.e., $A = \{\text{Do nothing}, \text{Repair } \% \text{ of track sections}\}$, with the percentages of track sections repaired being defined according to the policy being tested. For the expected costs equation, three different costs are considered: cost of repairing % of track sections, unavailability costs and cost of unplanned maintenance. It is assumed a fixed repair cost per track segment, and the cost of unplanned maintenance is also fixed, but 5 times higher than repairing costs in a preventive fashion. Unavailability costs vary according to a step function: since the track portion constitutes a single-track line, repairing implies closing the track and, thus, unavailability that is proportional to the number of segments being repaired. Hence, a step function with increasing costs as more sections are added is used to account for unavailability costs.

For the probability transition matrix, which varies according to the policy and H, some assumptions were made. First, the most conservative assumption is that the track can only move to state j if it is on state $j-1$, except for returns to state 1. This means that $p_{ij} > 0$, only if $j = i+1$ or $j = i$. For example, this means that if 20% of the track sections are degraded, then it is only possible to have the same amount or an increase to 30% of sections degraded in the next time period.

Due to Kriging's poor performance when extrapolating for many periods in time, the transition probabilities were obtained, instead, through a logistic regression model. In this model, it is assumed that the in-control design state corresponds to the overall mean from Kriging's model. Then, after the first time epoch of 6 months ($t = 1$), the probability that a state will exceed H is computed. Hence, the 49 track sections are modelled as independent Bernoulli Random Variables having parameter equal to the probability of exceeding H in in t time periods (π_t). This is be done by taking, from the database, all track sections whose twist is approximately equal to the overall mean and then considering the value one period afterwards: the ones exceeding H are marked with 1, 0 otherwise. For low thresholds H, this probability (π) turned out to be very high.

Since track degradation is considered in percentages which increase by 10%, moving

from one state to another implies in 10% more (or at least 5) sections degraded. The transition probabilities take this into consideration by using logistic regression to estimate the probability of a track section exceeding H in t time periods (π_t) and then using order statistics to get the probabilities of at least $t \times 5$ sections being degraded. For example, p_{12}, which is the probability of moving from a scenario where none of the track sections are degraded to a scenario where 10% (or 5) of track sections are degraded, is obtained by computing the probability of at least 5 Bernoulli(π) random variables (corresponding to 10% of track degraded) to be equal to 1, or equivalently, given $i = 45$ (number of sections which are not degraded) and $n = 49$ (total of sections):

$$\Pr(X_{45:49} = 1) = \Pr(\text{at least } 49 - 45 + 1 \ X_r\text{'s are} 1) = \sum_{r=49-45+1}^{49} \binom{49}{45} \pi_1^r (1 - \pi_1)^{49-r} \tag{4.14}$$

Same reasoning can be applied to get the other transition probabilities, under the assumptions above stated. For example, p_{23} will be obtained from to the logistic regression model to estimate the probability of a state that only went out-of-control after 2 time epochs ($t = 2$). This means that, for $t = 2$, another parameter π_2 is calculated from the Logistic Regression model and the transition probability p_{23} will correspond to the probability of at least 10 (or 20% out-of-control sections) Bernoulli(π) random variables to be equal to 1.

As expected, even considering that the parameter π_t increases after each time epoch t (because twist degrades over time), the probability that $t \times 5$ sections will be degraded decreases (as it can be inferred from Equation 4.14). At the same time, unplanned maintenance probabilities rise, and they are taken as the complement of the probabilities obtained with the logistic regression and order statistics approach, hence they are computed as $1 - \Pr(X_{(49-t\times5):49} = 1)$, where the last term is the order probability. Unplanned, or corrective maintenance actions, make the transition from the current state to state 1 happen, incurring in unplanned maintenance costs. The unplanned maintenance probabilities are

taken as the complement of the order statistics probabilities as mentioned above, except for states 1 and for the state for which the maintenance policy recommends a preventive return to state 1, which results in a mandatory return to the state where no track sections are degraded. For instance, if maintenance policy being tested says that track is repaired when 70% of sections are out-of-control, then at state 8 there is a mandatory return to state 1, with probability 1, and the probability of unplanned maintenance is 0, as the repair is preventive.

Finally, it is important to understand how each one of the nine different maintenance policies works (for each H). In this case, the i^{th} maintenance policy is related to making the repair as soon as the system gets to state i, for $i = 3, \cdots, 11$, with $i = 3$ corresponding to the policy of repairing as soon as 20% of the track sections get degraded over H and $i = 11$ corresponding to only making repairs when all track sections are degraded over H. For the i^{th} maintenance policy, $p_{i1} = 1$ (i.e. the Markov Chain is limited to i states).

The expected costs equation for a given H under i^{th} maintenance policy (considering m as the number of allowed track sections to be out-of-control, k as the number of track sections out-of-control in state j and unitary costs $c_r = c_{\text{repair}}$, $c_u = c_{\text{unplanned}}$ and $c_a = c_{\text{unavailability}}$ at time t is:

$$\mathrm{E}[c(t)] = c_r \cdot m \cdot p_{i1}(t) + c_a \cdot m \cdot p_{i1}(t) + \sum_j c_u \cdot k \cdot p_{j1}(t) + \sum_j c_a \cdot k \cdot p_{j1}(t) + \mathrm{E}[c(t-1)] \quad (4.15)$$

where $p_{i1}(t)$ is the probability of returning to state 1 at time t due to the maintenance policy i, and $p_{j1}(t)$ is the probability of returning to state 1 at time t due to the unplanned maintenance.

The unitary costs here were all assumed some fixed values. Hence, the results for 22 time epochs (or 11 years) will be showed in terms of percentages of the worst policy (highest expected costs) for each H in Table 4.2:

Table 4.2: Differences in % of expected costs

Policy	Threshold Value H			
(% out-of-control)	3	3.5	4	4.5
100%	62.7%	69.3%	73.1%	76.7%
90%	62.8%	69.7%	73.8%	76.9%
80%	64.1%	71.5%	75.5%	77.6%
70%	69.6%	76.1%	79.5%	79.6%
60%	80.7%	84.1%	86.3%	84.1%
50%	100%	96.9%	96.7%	93.4%
40%	99.9%	100%	100%	100%
30%	80.3%	86.7%	88.1%	93.1%
20%	60.3%	73.1%	76.3%	85.0%

Table 4.2 makes evident that the least cost options are either waiting for all track sections to be out-of-control or just doing maintenance as soon as 20% of sections are out-of-control. For example, considering 3 standard deviations for the threshold value H, the best policy is to do maintenance as soon as 20% of track sections go out-of-control, and this policy, over 11 years, could result in almost 40% savings for the railway company. In general, as can be seen from Table 4.2, the worst policies are those associated with repairs when about 50% of the track sections are out-of-control.

4.5 Conclusions and Further Research

This chapter proposed taking advantage of spatiotemporal data to build two Kriging models based on the Gaussian Correlation function to make predictions in the time domain for a transformation of the track's twist, which is one of the statistics related to track monitoring that can be used to measure track geometry degradation. Results for one time period extrapolation were satisfactory, although no maintenance work was considered. The Limit Kriging approach was preferred by its potential to model peaks and, therefore, indicate possible maintenance faults.

A commonly used charting method in SPC, the CUSUM Control Chart, with a slight correction to make it suitable for autocorrelated data based on the correlation function in Kriging, was then proposed and out-of-control points for different twist standard deviations

allowances (CUSUM's threshold H) were computed.

Lastly, it was desired to test different maintenance policies based on different percentages of track sections to repair and different thresholds H to have an idea of costs associated with each. Some assumptions were made to get probabilities of track sections being out-of-control at each epoch. These policies were compared in a simple MDP model, with all costs assumed. The results revealed the best maintenance policies as those in the boundaries, i.e. either performing maintenance when only 20% track sections are out-of-control or waiting for 90% or more to be out-of-control.

In terms of future research, this study has revealed many opportunities. The first one is related to the monitoring statistics: although twist was used as the main track parameter for this study, it is often used in the literature as a safety parameter. Other parameters, such as longitudinal level or a combination of parameters, may be more representative of the track's actual degradation state for the maintenance planning from the perspective of immediate action cost-optimization.

The second point is related to the modeling of the repair action in the Kriging model, where estimating the effects of the various maintenance actions (e.g. tamping, sleeper changes, ballast maintenance, etc.) and their effectiveness after a number of them have been performed is desired for an accurate representation of the actual repair process.

Finally, a better approach for getting the optimized strategy is also desired. Many of the assumptions adopted to build the MDP model are too conservative, especially the ones related to the derivation of the transition probability matrices, e.g. independent among different track sections. Therefore, an approach that can better incorporate the Kriging's predictions, which model the correlation among different track sections, is expected to improve the model.

CHAPTER 5
STATISTICAL REPRESENTATION OF RAILWAY TRACK IRREGULARITIES USING WAVELETS

5.1 Introduction

Inspection data in the form of onboard measurements collected by track recording vehicles (TRV) that run the railway track is often used for data analysis and maintenance planning purposes. This waveform type of data usually has high resolution and is sampled at very small intervals along the track to guarantee almost continuous monitoring. In terms of track geometry irregularities, they can be classified into short wavelength and long wavelength defects, with the former being commonly associated with irregularities that can potentially generate more vibration on axles and wheels and the latter to effects on the comfort of passengers [8]. This classification based on wavelength is also useful from the standpoint of track defects repair, where two main groups can be identified: i) those coming from the loss of track vertical geometry, whose wavelengths are usually longer than 2 m, and ii) those related to rail corrugation or isolated rail defects (e.g. squats, spalling), in which a loss of material is produced at the top of the heads of the rails while the track keeps its vertical alignment [117].

A train running on a track can be modeled as a system of masses, springs and dampers connected to each other, so that any significant change in any of the elements will affect the rest of the system. More specifically, the train can be modeled as a three masses system (the car body, the bogie frame and the wheelset) and the track as a two masses system (the rail and the sleeper), where the interaction between train and track happens in the wheel-rail contact [117]. In general, any vehicle passing by the track will generate vibration. However, different and complex deformation patterns acting in the track's structure make

the task of identifying and predicting the behavior of the various track's vibrational modes challenging, and it is often the case that multiple modes are found propagating at the same frequency [118]. In addition, for a general inspection of the railway track, there is a compromise between the resolution and amount of recorded data, which may cause some track defects whose excitation frequency is higher than the sampling frequency not to be detected. Identifying and avoiding some high-frequency defects, such as railhead roughness and very short railhead corrugation, is important as they produce a resonance effect in the train or in the track, which amplifies the train movements and increases dynamic loads and material wear, also increasing the risk of derailment [117].

Most railway infrastructure managers (IMs) assess the track quality either by comparing isolated track defects with track quality indicators, usually dependent on dispersion measures (e.g. standard deviations) of track irregularities [88]. In fact, this is the approach recommended in some railway standards, such as UIC 518 [12] and EN 13848 [1]. Although this approach has been widely used, it has some drawbacks. The first one is the averaging effect caused by considering long maintenance sections (MAINS), as highlighted by Esveld [9]. The consideration of long maintenance sections, usually 200 m, is often employed to justify the assumption of independence among the various sections, as the serial correlation approaches zero as the length of the track section increases. Second, such point-wise comparisons may be flawed since they ignore whatever has happened to the track up to that point. Li et al. [88] mention the importance of considering dynamic responses at the wheel-rail (or train-track) interface at different train speeds and loads. They claim that short-wavelength defects, such as rail welds, dipped joints and hanging sleepers are difficult to detect by evaluating only the measured track data and illustrate with an example where geometric irregularities of short wavelengths ($\lambda = 0.5 - 3$ m) generate high dynamic wheel-rail forces, although the irregularity amplitudes are below the limit values specified in standards. In a recent study, Balouchi et al. [119] revealed that a new sensor installed in the train vehicles combined with proper signal processing and fault detection

methods was able to capture features in the track for which the vehicle was responding to, although conventional track maintenance standards did not point out any potential alert limit extrapolation.

In light of the above paragraph, monitoring track geometric condition by considering only isolated track defects and corresponding track quality indexes may potentially result in ignoring the resonance effect in the train or in the track, compromising the adequate treatment of vehicle response to the various irregularities. The consideration of long track sections (e.g. 200 m) or point-wise defects does not provide a good location-frequency resolution, and, hence, the shape and frequency content of defects are not properly evaluated by the common methods, motivating the use of Wavelet Analysis.

Wavelet Analysis (WA) is an extension of the Fourier Analysis (FA) which provides a better location-frequency (or time-frequency) resolution. WA processes data by using a set of basis functions with finite energy, which is a good alternative for non-stationary or transient signals, allowing the detection of irregular events, very common in fault diagnostics [120]. WA multiresolution capability makes it a powerful joint time-frequency analysis technique, and it has gained increasing attention for the analysis of multimode and dispersive signals [118]. This chapter expands the scope of chapter 4, by including some other relevant track geometry parameters and considering the irregularities signals in the wavelet domain. It opens up a recent debate topic within track maintenance which refers to the use of predefined standard alert limits and presents an alternative assessment method based on using wavelets at different scales. In particular, the proposed method combines WA and vehicle dynamics simulations to identify track faults by using the traditional Nadal's safety criterion for derailment, known as Y/Q.

5.2 Related Work

Much of the related work associated with track fault prediction and track geometry maintenance scheduling was covered in chapter 4. This chapter provides related work

associated with the use of WA, which is the methodology adopted herein.

Koziol [121] highlights the necessity of new methods for the parametrical analysis of dynamic systems such as the railway track and mentions some of the advantages of having mathematical models along with effective procedures to replace commonly used numerical computations involved in the analysis of track's dynamic behavior. He presents an experimental validation of the semi-analytical wavelet-based technique to solve nonlinear models associated with railway track subject to moving loads and validates the approach using a single-layer model, where only the rail layer is considered. The model is solved by using Adomian's decomposition combined with the wavelet-based approximation of the Fourier integrals, for the case where support stiffness is nonlinear. The author mentions that classical methods of computation of the Fourier transform and inverse Fourier transform become ineffective due to the complex form of the integrands combined with the Adomian polynomials formulation, but the novel method using wavelet approximation through coiflet filters reflects the dynamic features of the system well enough. This is an important contribution as the modeling of nonlinear properties and stochastic variations of physical characteristics in the railway track context is still considered an open problem in the literature, due to lack of appropriate tools to obtain analytical solutions.

Zhiping et al. [122] apply WA to decompose a track irregularity signal and use Fast Fourier Transform (FFT) to get the power spectral density (PSD) curves of the different decomposition levels. By doing so, they distinguish different wavelength content which are captured in different decomposition levels and associate these to possible defects or error. For example, at decomposition level 7, the detail coefficients reflect $33 - 50$ m wavelength irregularities, which may be caused by bridge creep and track adjustment error.

Salvador et al. [117] use axle box accelerations, i.e. accelerations measured at the ends of the wheelset's axle, to present an analysis tool, by means of an appropriate tuning of the spectrogram defining parameters, capable of showing track vibration modes and defects, although heavily dependent on the interpretation and knowledge of the maintenance

analyst. They further compare their surveying approach to WA, and conclude that WA does not properly show vibration modes with frequency higher than 100 Hz, although detection of track singularities may be improved with respect to their method using the spectrogram. However, they also point out that WA performance is affected by the choice of mother wavelet and the resolution level, leaving this research topic as possible further investigation.

Some authors have used WA for fault diagnosis in the railway track. For example, Xu et al. [123] build a model for predicting track portions with deteriorated wheel-rail forces by combining wavelet transform and Wigner-Ville distribution for characterizing time-frequency characteristics of track irregularities and a three-dimensional nonlinear model for describing vehicle-track interaction. Caprioli et al. [124] show the promising aspects of the wavelet approach (both continuous and discrete) in comparison to Fourier Analysis for fault detection in rails from measurements carried out on axle box accelerations running on the line. They propose the creation of an ad-hoc wavelet, suitable to get out a particular defect ”stamp” and by using wavelet packets (a natural extension of DWT) they show the method's ability to detect short pitch corrugation, changes in track sub-structure and local defects. Toliyat et al. [125] also use wavelet packets as the main approach for the detection of defects in rail, and they show the effectiveness of the method by comparing the deviation of wavelet coefficients in the ”healthy” rail from the rail containing defects. Shah et al. [126] use Canny edge detection for fast screening and identification of track damage and further decompose the images where damage was identified with 2D discrete wavelet transform. Most of the edge background details are classified into the high-frequency sub-band by the wavelet transformation whereas noises and approximated signal are classified into the low-frequency sub-band, so that the transformation is effective for detecting the severity of damage and the surface area that is most affected by it. Hopkins and Taheri [127] propose a monitoring algorithm that could be used with a portable defect detection system consisting of accelerometers mounted to the side frame of a bogie placed on an in-service

train vehicle. The method uses discrete wavelet transform as the basis for diagnosing effects under this limited-data scenario and further calculates their degree of irregularity by utilizing the Lipschitz exponent. A vertical acceleration test signal is overlaid with the locations of defects, which are known. At each scale of the DWT, a threshold level obtained by multiplying the average of the modulus maxima by 1.5 is calculated, and all coefficients at that scale below the threshold are set to zero. The values of the wavelet coefficients that made it past the threshold were shown to be related to wheel and rail defects, meaning that the algorithm was able to locate all important defects: wheel flats were found in the first two scales and cracks showed up in the first three scales (high-frequency content), all three sets of corrugation appeared in the first five scales, which agrees with the fact that all instances of corrugation have a range of varying wavelength and a broader range of frequency content.

As in some examples above, most applications focus on using the similarity of the signal under study and some mother wavelet. Still, the method of choosing a proper mother wavelet has primarily been that of trial and error and fault detection is performed by contrasting changes in the wavelet coefficients from the healthy rail signal to the defective one [128]. In a different application, Andrade et al. [129] use the wavestrapping technique [130] to reconstruct signals with the same statistical properties of an initial signal and to estimate extreme values in train aerodynamics, comparing the novel approach to the one presented in the standards.

Conventionally used track inspection vehicle data is costly and cannot be acquired as frequently due to operational schedules. Therefore, some authors rely on data from sensors installed on operational trains or data acquired inside laboratories. Kojima et al. [131] use data from a simple accelerometer installed on the board of a train and multi-resolution analysis (MRA) to show that the proposed method is able to detect corrugation by comparing the MRA of a straight track section with no corrugation and a curved one with corrugation. Zhang et al. [91] employ WA and Shannon entropy to detect

rail defects using data acquired from sensors in a laboratory rail-wheel test rig. The optimal wavelet is selected based on the appropriate decomposition level and Energy-to-Shannon entropy ratio. The study further encompasses the length of the time window in order to suppress noise effects and ensure appropriate time resolution. The monitoring method used in the study is acoustic emission (AE), which is suitable for investigating the dynamic behavior of materials and structure, making it ideal for on-line continuous monitoring. However, most of the literature on AE to detect track defects only considers trains at low speeds. The approach in [91] also deals with high speeds cases (124 km/h), in which most defect signals are submerged in noise. WA is used to effectively extract the defect features, for its ability to detect local variations of non-stationary signals (such as those coming from AE) accurately. Scalea and McNamara [118] use WA to investigate the behavior of high-frequency longitudinal and lateral transient vibrations propagating in railway tracks in a 4.6 m laboratory track. They show the supremacy of WA with respect to short-time Fourier transform (STFT) and some other methods that also optimize the time-frequency resolution, such as pseudo-Wigner-Ville distribution (PWVD), and conclude that the technique is well-suited for extracting both the group velocity dispersion and the frequency-dependent attenuation of the various track vibrational modes. Gomez et al. [132] use axle box measurements of a full bogie installed on a rig to propose a real-time condition-monitoring technique based on vibration analysis by means of the Wavelet Packet Transform (WPT) energy, combined with a Support Vector Machine (SVM) diagnosis model. Cracks were induced in the axles and a comparison with the healthy axle was made, so that any changes observed in the signals for each wheelset could only be attributed to the appearance of the defect. Although the number of false alarms for the SVM linear model was high in the test data when only a limited number of wheelsets were used, the authors claim that a diagnosis model built with more data can increase the model's accuracy.

In more recent work, Balouchi et. al [119] present a monitoring approach based on a low-cost sensor which makes use of the existing on-board GSM-R cab radio present in

the majority of trains operating in the UK to monitor the vibration in the vehicle in three axes, allowing detection of certain track features. The authors highlight that many train companies worldwide obtain monitoring data through specialized measurement vehicles that run the track, which corresponds to the type of data investigated in the case study covered in chapter 4 and also in this chapter. However, to increase coverage and frequency of track geometry monitoring, many companies are adding unattended track geometry measurement systems (UGMS) in the passenger vehicles, during normal train operation. Although this practice increases the frequency of data available, it has a major drawback: the vehicle air suspension, that is designed to improve passenger comfort, eliminates the high frequencies components of the signal, although the amplitude of a significant impact is still transferred into the car body, potentially causing track faults. The novel sensor presented in [119], as indicated by the results of several vehicle dynamics simulations, is able to capture the location and severity of irregularities at the vehicle-track interface and also to identify the type of track asset (e.g. switches and crossing, structure or plain-line track), overcoming the major drawback from previous UGMS. The signal processing technique adopted by the authors was the continuous Wavelet transform (CWT), which has good time-frequency resolution, allowing a ”windowed” signal estimation sized according to the frequency and, therefore, preserving the time of the detected irregularity. One of the important conclusions of the work in [119] was that filtering the acceleration response at different frequency ranges allows the system to differentiate between the various types of potential track faults, where, for example, low-frequency features could be related to voided sleepers and high-frequency features to defects, such as corrugation and wheel flats. The outlined frequency analysis was very important from the perspective of isolating the signals which were indicative of track faults, hence crucial in the development of the detection algorithm, being further assessed through vehicle dynamics simulations.

Recent literature also highlights a research gap related to the challenge of finding the statistical correlations between track geometry quality and vehicle dynamics, given varying

relationships even on areas of identical track quality [133]. Resampling the same wavelet coefficients multiple times and measuring the effects on the response allows, at least in theory, to study which configurations or combinations of coefficients can potentially impact the response and cause a future failure. However, relating the amplitude of a signal with abnormal track features is not an easy task because it requires a complex model that tackles the correlation between the various parameters and how they directly impact the track. An alternative way may be to run different vehicle dynamics simulations which differ from each other only by some few irregularities coefficients. This methodology of ”keep everything the same, except for some few coefficients” and then compare the results is, hence, an alternative to potentially overcome some of the problems related to the complex nature of the railway track that was highlighted above.

5.2.1 Main Contributions

In this chapter, one of the research opportunities outlined in [129] and [119] is further explored, namely the track response, and consequently track fault prediction, subject to variations in the inputs in forms of different wavelengths and amplitudes. In particular, WA is used to study and reconstruct different track geometry irregularity signals. Then, an approach similar to wavestrapping [130] is adopted, but instead of sampling from the set of coefficients that corresponds to the original data transformed into the wavelet domain, a signal is resampled and reconstructed based on some pre-defined decomposition levels of a chosen wavelet. These decomposition levels can be roughly understood as different grades of wavelet coefficients vectors, each of which includes different time window and frequency window (frequency decreases as the decomposition level increases). By reconstructing the different irregularities in this way, it is desired to find the frequencies that are more likely to generate successive excitations of the car body, leading to maximized force transmissions between vehicle and rail. In other words, the first contribution of this chapter is to identify which wavelets are more present in typical railway track irregularities,

and to investigate whether the presence of some high amplitude wavelet coefficients in certain frequencies can be associated with higher vertical or lateral forces in the wheel-rail contact. The last step is accomplished by reconstructing the different irregularities signals using wavelets coefficients in various decomposition levels and studying their impact on Nadal's safety criterion Y/Q (a critical quantity for derailment safety assessments) through vehicle dynamics simulations. The presence of certain wavelets at different decomposition levels allows identifying wavelets that are more prejudicial in terms of the safety criterion.

For the detection of track faults, experimental studies are very scarce due to the multimode (multiple modes propagating at the same frequency) and dispersive (propagation velocities depending on the frequency) characteristics of the waveform data [118]. Both Fourier transform and Wavelet transform have been used in the literature. However, Fourier transform has worse time resolution, meaning that important information for determining the location of faults may be lost. Even when a windowed Fourier transform is used, it still depends on the size of the window applied and, hence, it may not be suitable for detecting unknown signals [131]. A joint time-frequency analysis technique such as Wavelet transform is an alternative to methods that use multiple, equally-spaced waveforms [118]. There are only a few applications in the railway field encompassing the study of the criteria for an adequate choice of wavelet. On the other hand, the selection of the wavelet plays an important role, since track defects appearance and vibration patterns strongly depend on this parameter [117]. No studies in the literature were found to relate the similarity of some particular geometric defects with the "shape" of a selected wavelet, so the other main contribution of the investigation in this chapter is to find evidence of wavelets and respective scales which can be associated to increased risk of derailment (in terms of Y/Q), even when the original irregularity signal measured at that point has not surpassed the immediate action limit (IAL) suggested in the conventional track maintenance standards, in this case, the European Standard EN 13848-5 [1].

The findings herein raise questions on the appropriateness of applying the same track

maintenance standards to all train operating companies and track irregularities based on point-wise comparisons with alert limits.

5.3 Preliminares

5.3.1 Wavelet Analysis

Wavelets are mathematical objects which have broad application in "time-scale" types of problems. The term is usually associated with a function $\psi \in L_2(\mathbb{R})$ such that the translations and dyadic dilations of ψ, $\psi_{jk}(x) = 2^j\psi\left(2^j x - k\right)$, $j, k \in \mathbb{Z}$, constitute an orthonormal basis of $L_2(\mathbb{R})$ [134]. In comparison to Fourier transformations (FT), although they behave similarly, wavelet transformations have the advantage of being able to describe the evolution of the spectral features of a signal as it evolves in time or space [124]. Because wavelets come in different shapes and are limited in time and frequency, decomposing a signal using wavelets can give a much better resolution than using FT, where the signal is decomposed into sinusoidal waves that are infinitely long. Therefore, the Fourier Transform extracts details from the signal frequency, but all information about the location of a particular frequency within the signal is lost [134]. The best that can be done with FT is to sample a range of time or space and find a range of frequencies existing over that amount of time or space, meaning that one has to evaluate the trade-off between knowing precisely the frequency or time of the signal, but not both. The theoretical justification lies in the Heisenberg uncertainty principle, which implies that a constant time window cannot maintain adequate resolution in the low-frequency and in the high-frequency ranges simultaneously [118]. Wavelets, on the other hand, are finite, can be shifted along time domain and be compressed (high frequencies) or stretched (low frequencies), which configures as a multiresolution capability that is broader in time for observing low frequencies and shorter in time for observing high frequencies [118]. Therefore, WA is an alternative to cope with the limitations of Heisenberg's uncertainty principle by combining both time and frequency domains, a notable advantage compared

to the FT.

Wavelet transforms can be either continuous or discrete. The discrete transformation (DWT) can be efficiently realized by decomposing the signal into approximation (low frequency) and detail (high frequency) coefficients, and after each level of decomposition only the approximation is decomposed into further level. Unlike the DWT, the continuous transformation (CWT) can operate at every scale [135].

Let $\psi(a,b)(x)$, $a \in \mathbb{R} \setminus \{0\}$, $b \in \mathbb{R}$ be a family of functions defined as translations in x (assigned by $b \in \mathbb{R}$) and re-scales (assigned by $a \in \mathbb{R} \setminus \{0\}$) of a single function $\psi(x) \in L_2(\mathbb{R})$, also called the wavelet function or "mother" wavelet as in Equation 5.1:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{x-b}{a} \right) \tag{5.1}$$

The Continuous Wavelet Transform (CWT) is obtained by convolving a signal with an infinite number of functions, generated by translating and re-scaling a certain mother wavelet $\psi(x)$. The CWT for a signal $f(x)$ is defined as a function of two variables, as shown in Equation 5.2:

$$\text{CWT}_f(a,b) = \langle f, \psi_{a,b} \rangle = \int f(x) \overline{\psi_{a,b}(x)} dx \tag{5.2}$$

In the resulting transform the dilation and translation parameters, a and b, respectively, vary continuously over $\mathbb{R} \setminus \{0\} \times \mathbb{R}$. The signal $f(x)$ is a function of one parameter and its CWT is a function of two. To make the transformation "less redundant", one can select discrete values of a and b and still have a transformation that is invertible: $a = 2^{-1}$, $b = k2^{-j}$ will produce a minimal basis, so that any coarser sampling will not give a unique inverse transformation; i.e., the original function will not be uniquely recoverable [134]. Such sampling, under some mild conditions, produces the aforementioned orthogonal basis $\{\psi_{jk}(x) = 2^j \psi \left(2^j x - k\right), j, k \in \mathbb{Z}\}$. The "mother" wavelet can be rewritten according to

Equation 5.3:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{x - 2^j \cdot k}{2^j}\right) \tag{5.3}$$

In CWT the output comes in the form of smoothly varying local frequency and scales. However, a more efficient way to extract the signal's interesting features and reconstruct the signal by using only a few important coefficients is to consider Discrete Wavelet Transforms (DWT). Since track monitoring data is often large and DWT is considerably faster than CWT, DWT is selected to analyze the signals. Mallat [136] proposed an efficient pyramid algorithm in which a repetitive application of high pass and low pass filters decomposes the original signal and allows to extracextracting the difference of information between successive resolutions.

The DWT uses the concept of multiresolution, which can be roughly understood as a decomposition of a signal on a grid of time and scale (frequency). In practice, only a limited number of scales and positions based on powers of two are used (dyadic scales and positions) [137]. The wavelet transform requires a pair of filters, one that computes the wavelet coefficients and the other that applies the scaling function. The scaling function ϕ, also called "father" wavelet, is auxiliary to the mother wavelet function (storing the approximations or the "remaining information"), and is defined as in Equation 5.4:

$$\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{x - 2^j \cdot k}{2^j}\right) \tag{5.4}$$

The scaling function is the complement of the wavelet, and just like the mother wavelet, the father wavelet can be translated and dilated over the signal. Following the above notation, and considering a signal $f(x)$, where x is taken to be a certain location of the track, $x = 0, 1, 2, \cdots, N-1$, and with $N = 2^j$, where N represents the total track length. Then, $j = 1, 2, \cdots, J_0$ indexes the scale $S_j = 2^j$ $\left(S_1 = 2, S_2 = 4, \cdots, S_{J_0} = 2^{J_0}\right)$ to which the wavelet has been dilated ($J_0 < J$ defines the maximum scale of the analysis)

and $k = 1, 2, 3, \cdots, K_j$ indexes the location in the space to which it has been translated, where $K_j = \frac{N}{2^j}$ implying $K_1 = \frac{N}{2^1}, K_2 = \frac{N}{2^2}, \cdots, K_{J_0} = \frac{N}{2^{J_0}}$. The wavelet coefficients $d_{j,k}$, associated with the mother wavelet, and the scaling (expansion) coefficients $a_{j.k}$, associated with the father wavelet, are defined in a linear signal decomposition as in Equation 5.5:

$$f(x) = \sum_{k=1}^{2^{J-J_0}} a_{J_0,k}\phi_{J_0,k}(x) + \sum_{j=1}^{J_0}\sum_{k=1}^{2^{J-j}} d_{j,k}\psi_{j,k}(x) \tag{5.5}$$

The above equation shows how a signal $f(x)$ can be decomposed as the sum of an approximate signal (expanded in terms of scaling functions), and of several detail signals expanded in terms of wavelets. The scale corresponding to J_0 represents the biggest scale of the wavelet. Calculation of the expansion coefficients may be performed relying on a two-channel filter bank downsampled by a factor 2. Hence, the DWT works by iteratively filtering the signal with a low-pass filter (father wavelet) and a high-pass filter (mother wavelet), downsampling the results by a factor of 2, so that the remaining coefficients are of the same dimension of the original signal and repeat these steps on the smaller low-passed signal a number of times (according to the level of decomposition chosen). When several two-channel filter banks are connected repeatedly at the output of a low-pass filter a tree dyadic structure is obtained: this is the conventional way to build up the DWT [124]. The DWT coefficients will have the same dimension as the original signal. These coefficients represent the energy level of the decomposed signal in each frequency band, with the first iteration having the high-frequency coefficients or fine-detailed features, moving to iterations linked to low frequency or coarse features.

The first aspect to be considered in DWT is the choice of the mother wavelet. This selection is a crucial step as different mother wavelets applied to the same signal may produce different results. Moreover, previous research has indicated that the choice of mother wavelet and scaling function is application-dependent [128]. Properties of the mother wavelet, such as orthogonality, support, symmetry and number of vanishing moments can be taken into account, for example orthogonal wavelets are not not redundant

and may be suitable for signal denoising [128]. However, more than one mother wavelet with the same properties may exist [135]. The similarity of the original signal and mother wavelet is often considered, and previous studies have compared similarity by either qualitative and quantitative methods. In the qualitative approach, for example, Ahadi et al. [138] compare a signal with two clear transient events at two different times. By visual inspection, they determine that the ”haar” mother wavelet is more suitable to model the transitions from high to low amplitudes and vice versa, relatively to the ”db8” mother wavelet. The comparison of spectrograms of the WT clearly shows a lack of localization of the ”db8” mother wavelet, as the zones indicative of large values of the WT are spread along the time axis, whereas the ”haar” spectrogram shows the two transient events well localized in the time axis. In their application of leak detection in water-filled plastic pipers through acoustic emission signals, they argue that only a transformation that is well localized in time enables detection of leakage, as the energy will be significantly larger compared to that of background noise (larger signal-to-noise ratio). As far as shape matching and qualitative approaches are concerned, it is generally difficult to accurately match the shape of a signal to that of a base wavelet through a visual comparison [139].

Quantitative approaches aim to find proper criteria that can be measured or quantified in order to select the most adequate mother wavelet for the analysis. For example, Toliyat et al. [125] use an algorithm called the Minimum Description Length (MDL) to optimize the efficiency of compression in the wavelet domain. The MDL criterion enables selection of appropriate mother wavelet and number of retained coefficients in a signal by evaluating the trade-offs between the number of retained coefficients and the error of signal reconstruction. Yang et al. [140] use the correlation coefficient criterion, which evaluates the similarity between the signal and the wavelet function. Tsui et al [141] perform wavelet basis selection by deriving closed-form solution of the information measure, symmetric divergence (also known as relative entropy), which is employed as a similarity measure between different classes of materials. The best mother wavelet, for classification tasks

presented in the study, is the one returning the largest divergence. Another example is given in the work of Rafiee and Tse [142], who propose computation of variances of wavelet coefficients as the main measure. The authors state that the more variance available, the better the ability of the mother wavelet to properly classify failures. He et al. [139] state that entropy can measure the features of uncertainty associated with the electrocardiogram (ECG) signals and, hence, they propose a novel comprehensive entropy criterion based on multiple criteria related to entropy and energy. They further experiment and validate the method on the basis of ECG signals of sixteen subjects selected from the MIT-BIH Arrhythmia Database.

In this chapter, the concept of entropy is also explored and it is used as a criterion for wavelet basis selection. Entropy is one of the most fundamental concepts in science [143] and is a measure that can be used to evaluate the uncertainty associated with a random variable, i.e., the expected value of information in a message [139]. The energy distribution of wavelet coefficients is quantitatively described by Shannon entropy [91], which is one of the most important metrics in information theory [139]. Higher energy concentration translates into lower entropy value, hence the minimization of entropy leads to maximization of energy.

Following Oliveira [143], entropy of a random variable X with probability density $p(x)$ is defined by Equation 5.6:

$$H(X) = -\int_{-\infty}^{\infty} p(x) \cdot \log p(x) dx \tag{5.6}$$

The base of the logarithm defines the information unit, and Shannon entropy has binary unit. Hence, for a continuous wavelet $\psi(\cdot)$, in terms of Shannon entropy, the time entropy, $H_t(\psi)$, and the frequency entropy, $H_f(\psi)$, can be written as in Equation 5.7 and

in Equation 5.8, respectively:

$$H_t(\psi) = -\int_{-\infty}^{\infty} \psi^2(t) \cdot \log_2 \psi^2(t) dt \tag{5.7}$$

$$H_f(\psi) = -\int_{-\infty}^{\infty} \frac{1}{2\pi} |\Psi(\omega)|^2 \cdot \log_2 \frac{1}{2\pi} |\Psi(\omega)|^2 d\omega \tag{5.8}$$

where $\psi(t)$ is a continuos wavelet and $\Psi(\omega)$ is the Fourier transform of $\psi(t)$. In this way, the probability density functions $p_t(\psi) = \psi^2(t)$ and $p_f(\psi) = \frac{1}{2\pi}|\Psi(\omega)|^2$ can be associated with a continuous wavelet $\psi(t)$.

The wavelets explored in this chapter, namely Daubechies, Symmlets and Coiflets, cannot be described by analytical expressions, except the Haar wavelet, which is a special case of the Daubechies. Instead, they are expressed via filter coefficients. For the purpose of this chapter, the Shannon entropy is calculated by considering signal data $\boldsymbol{X}_{p \times N}$, where $p \in \mathcal{L} = \{1, \cdots, p\}$ is the total number of input signals and $N \in \mathcal{M} = \{1, \cdots, N\}$ is the length of each input vector. Hence, $\boldsymbol{X}_l$ represents all observations of the signal at index l. The wavelet transform can be then represented by $\mathcal{W}(\boldsymbol{X})$ and the vector $\boldsymbol{w} = [\boldsymbol{W}_1^T, \cdots, \boldsymbol{W}_p^T]$ of length $N \times p$ corresponds to the concatenated output in form of wavelet coefficients. For one signal, the unidimensional vector $\boldsymbol{w}$ can be computed according to Equation 5.9:

$$H(\boldsymbol{w}) = -\sum_{i=1}^{N} w_i^2 \cdot \log w_i^2 \tag{5.9}$$

From Equation 5.9, it is clear that Shannon entropy is calculated on the squared values of wavelet coefficients. The entropy obtained is further normalized by the coefficient values and, in case $\boldsymbol{w}$ is not unidimensional, the total number of signals as well.

5.4 Application of WA and Assessment of Y/Q with Vehicle Dynamics Simulations

5.4.1 Track Irregularities

Track geometry deteriorates under the influence of dynamic track loads. These loads cause stresses and elastic displacement and, depending on the total stress level, also permanent deformations [9]. In chapter 4, the deterioration of track geometry was quantified by using the track twist. In this chapter, four other parameters are explored: longitudinal level, lateral alignment, cross-level, and gauge. These four parameters are listed in the European Standard EN 13848-5 [1] as predominant influence parameters on vehicle response in terms of Y/Q. Following Soleimanmeigouni et al. [8], longitudinal level is the track geometry of track centreline projected onto a longitudinal vertical plane. Lateral alignment is the track geometry of track centreline projected onto a longitudinal horizontal plane. Cross-level is the difference in the height of the adjacent running tables computed from the angle between the running surface and a horizontal reference plane. Gauge is the distance between the gauge profiles of two adjacent rails at a given location below the running surface.

The random nature of these geometric irregularities, wear on the rail profile, variations in track stiffness, and track structural issues in addition to deterioration of systems used in the railway vehicles are sources of the stochastic nature of track–rail interaction [10].

The track portion considered in this study is a 2048 m line (from 109.3575 km to 111.4055 km, consisting of UIC 60 kg/m rail profiles laid out in Iberian gauge (1.668 m), with curvature profile as shown in Figure 5.1.

The four raw signals corresponding to the measured track geometry components of longitudinal level, lateral alignment, cross-level and gauge, all obtained through a track inspection vehicle, are showed in Figure 5.2. These measurements come from a inspection vehicle that uses a contactless geometry measuring system associated with a GPS location device.

Each of the four measured signals in Figure 5.2 is a vector with length equal to

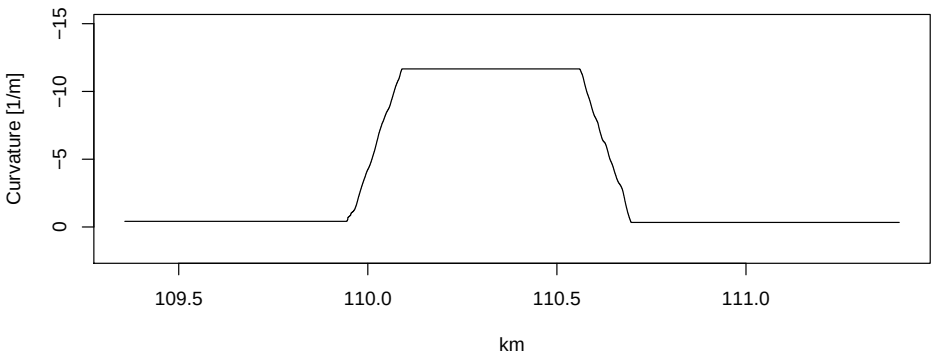

Figure 5.1: Curvature profile of track

$8192 = 2^{13}$, meaning that data (the 2048 m track segment) was sampled at 0.25 m intervals.

5.4.2 Alert Limits for Track Geometry Parameters

Periodic inspections through special vehicles that run the track are widely used as an instrument to measure the different track parameters and identify possible defects, by means of signal digital processing techniques [13, 14, 15, 16]. Often these signal processing techniques include the utilization of filters that focus on some wavelength band and eliminate wavelengths outside that band [14]. UIC 518 [12] and EN 13848 [1] are examples of standards that include alert limits for geometric parameters for different wavelength ranges. Since UIC 518 is more intended for testing the approval of railway vehicles, rather than track geometry maintenance purposes [14], the European Standard EN 13848 will be used in the remainder of this chapter.

EN 13848 puts forward a series of recommendations for the main track geometry quality indicators. It has been published in six parts. Part 5, which explores the geometric quality levels, is the main subject of interest of this subsection. EN 13848-5 recommends three main indicators to assess track geometric quality: 1) extreme values of isolated defects; 2) standard deviation over a defined length, typically 200 m; 3) mean value. It further defines three monitoring levels: (1) the immediate action limit (IAL), which refers

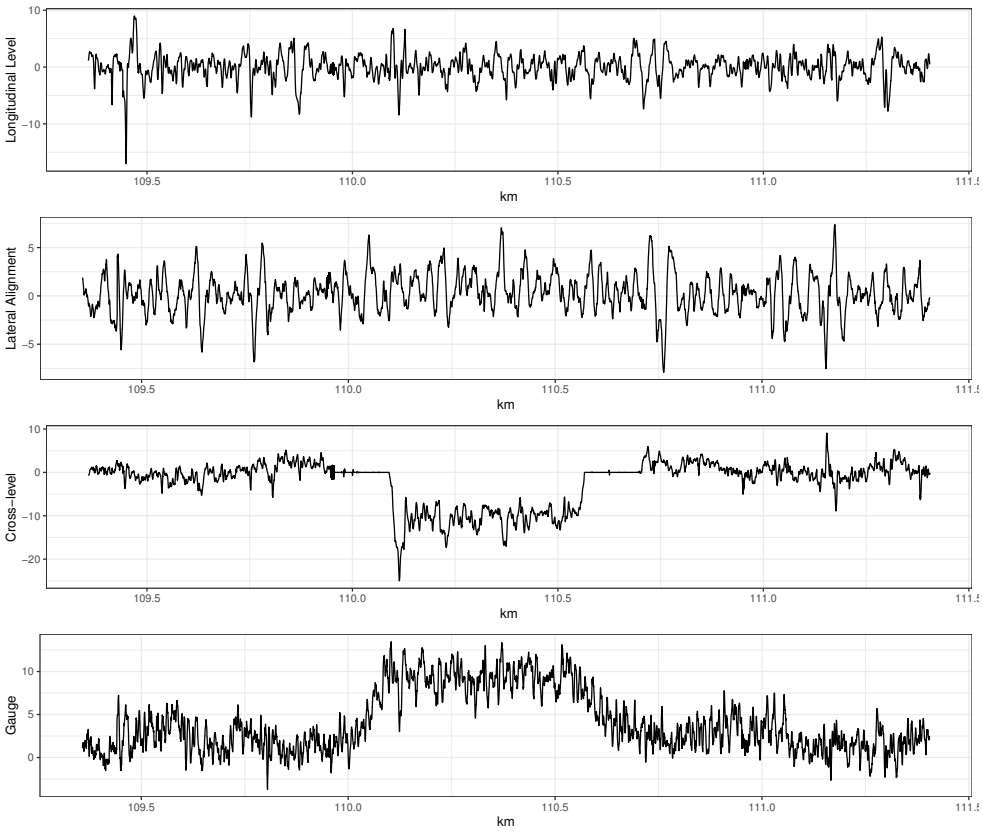

Figure 5.2: Track irregularities signals

to the value that, if exceeded, requires taking measures, such as speed restrictions or immediate correction of track geometry, to reduce the risk of derailment to an acceptable level; (2) the intervention limit (IL), which refers to the value that, if exceeded, requires corrective maintenance in order that the immediate action limit shall not be reached before the next inspection; and (3) the alert limit (AL), which refers to the value that, if exceeded, requires that track geometry condition is analyzed and considered in the regularly planned maintenance operation [1]. From these three items, only the IAL is normative, whereas IL and AL are given as recommendations, reflecting common practice among IMs.

The IAL values will be used as a reference for threshold limits in this study since they are normative. For some parameters, namely the longitudinal value and lateral alignment, IALs are given as a function of the wavelength range. EN 13848-5 includes three of such ranges, being them: $D1$ which ranges from $3 - 25$m, $D2$ which ranges from $25 - 70$m

and $D3$ which ranges from $70 - 150$m for longitudinal level and $70 - 200$m for lateral alignment. The standard also mentions the possibility of modifying $D1$ to a lower limit of 1 m, in order to include short-wavelength defects, but no changes in terms of IAL are considered. In addition, it mentions that $D3$ is used to detect long-wavelength defects, usually only considered for line speeds greater than 230 km/h. Hence, the $D3$ wavelength range is not taken into account in terms of IAL, since it is not directly linked with safety, but rather vehicle ride quality (or comfort of passengers).

Table 5.1 shows the immediate action limits (IAL) established in EN 13848-5 for different train speeds for the parameters under study. The standard does not provide IAL for cross-level because the risk associated with a cross-level defect is tied to twist and cant deficiency [1]. The values in Table 5.1 are considered to be mean to peak values for longitudinal level and lateral alignment, and nominal 1.668 m to peak for the gauge. In practice, however, the mean will be close to zero and zero to peak values may be used instead.

Table 5.1: Immediate action limits for different train speeds and parameters according to European Standard EN 13848-5 [1]

Train Speed s	Longitudinal Level (mm)		Lateral Alignment (mm)		Gauge (mm)	
(km/h)	$D1$	$D2$	$D1$	$D2$	Minimum	Maximum
$s \leq 80$	28	N/A	22	N/A	-11	$+35$
$80 < s \leq 120$	26	N/A	17	N/A	-11	$+35$
$120 < s \leq 160$	23	N/A	14	N/A	-10	$+35$
$160 < s \leq 230$	20	33	12	24	-7	$+28$
$230 < s \leq 300$	16	28	10	20	-5	$+28$

5.4.3 Y/Q Ratio Criterion in Derailment Safety Assessment

Derailment occurs when a vehicle runs off its rails. It may be caused by a number of factors, having as an immediate consequence the temporary disruption of the train operations and may potentially involve serious safety impacts (minor and major injuries, or even fatalities). Following Mohammadzadeh et al. [10], there are two types of derailment: (i) sudden derailment, caused by the wheelset jumping the rails and (ii) flange climb derailment,

caused by a wheel gradually climbing to the top of the railhead and then running over the rail. Regarding the latter, the wheel climbing to the top of the railhead is usually related to forces caused by track irregularities, and, hence, will be the focus of this study.

In terms of derailment, many criteria (see [10]) have been developed to provide railway companies with safe, and yet most times conservative, limits for operation. In this chapter, Nadal's criterion [144], which was developed in 1908 but is still widely used in derailment researches [10], especially those related to flange climbing, is considered. The criterion is based on a single wheel lateral to vertical force ratio (Y/Q), and the European Standard EN 13463 [145] defines the limits for safety operations against derailments, based on a track with a curve radius of $R \geq 250\,\text{m}$ and Y/Q per wheel previously filtered with a simple sliding mean over 2 m of track, to be according to Equation 5.10:

$$\frac{Y}{Q} < 0.8 \tag{5.10}$$

If the inequality in Equation 5.10 has been established, derailment due to the flange climbing will not occur.

5.4.4 Experimental Setup

This chapter assumes that WA is a method of investigation for the detection and evaluation of possible geometric track defects which have an impact on the Y/Q criterion, considering the different and complex deformation patterns acting in the track's structure. Wavelet transforms, as mentioned earlier, are limited in time and frequency, have a much better resolution than FT, making them an attractive alternative for fault diagnosis, as they can give better information about the (spatial) location of occurrence of the fault. They can detect discontinuities in the signal that occur in short intervals of time (highly localized) by using the transform at a fine scale or to capture several low-frequency cycles occurring over a broad interval of time by using a transform at a coarser scale [128].

On top of that, the investigation in this chapter aims at finding evidence of wavelets

and respective scales which can be associated to increased Y/Q, even when the original irregularity signal measured at that point has not surpassed the immediate action limit (IAL) suggested in the conventional track maintenance standards, in this case, the European Standard EN 13848-5 [1]. This agrees with the assumed hypothesis that the various defects have a range of varying wavelength and frequency content, and their observance in some specific "shapes" in the original signal could be an indicator of track faults that are potentially more penalizing in terms of Y/Q. These findings raise questions on the appropriateness of applying the same track maintenance standards to all train operating companies and track irregularities.

The experimental design can be further divided into three phases, where all reconstructed signals are simulated and compared to the base scenario (measured signals) in terms of Y/Q, as follows:

- Phase 1: reconstruction of all four irregularities signals based on 50%, 100% and 200% increase in all wavelet coefficients, resulting in 3 simulations for each wavelet tested;

- Phase 2: for each one of the 13 decomposition levels, reconstruction of all four irregularities signals one decomposition level at a time, while keeping the other scales the same, based on 50%, 100% and 200% increase in the wavelet coefficients, resulting in $3 \times 13 = 39$ simulations for each wavelet tested;

- Phase 3: for each one of the 13 decomposition levels, reconstruction of only one irregularity signal and one decomposition level at a time, while keeping the other irregularities and scales the same, based on 50%, 100% and 200% increase in the wavelet coefficients, resulting in $3 \times 13 \times 4 = 156$ simulations for each wavelet tested.

The total number of simulations amounts to 198 if all three phases above are considered. This number assumes the choice of single mother wavelet. Hence, if more wavelets are

tested, say k, then the amount of simulations totals $k \times 198$. Since there are many choices of wavelets available, a criterion has to be established to choose an appropriate one. Shannon entropy criterion is used to select an initial candidate for each of the irregularity signals from a pool of 26 wavelets. Then, another wavelet, which will have its performance compared to the minimum entropy wavelet, is selected based on a simulation experiment. This set of simulations resembles phase 2 simulations and it aims to find a mother wavelet, among the 26 candidate wavelets, that has the highest associated response (in terms of Y/Q) when reconstructed, in other words, the "most penalizing" mother wavelet in terms of the potential of generating Y/Q peaks. The reconstruction of the signals for this new set of simulations is performed according to the pseudo-code below:

Algorithm 1: Pseudo-code to find "most penalizing" mother wavelet

```
for each wavelet i = 1 to 26 do
    for each decomposition level j = 1 to 8 do
        i. non-alternating signs scenario: make all wavelet coefficients equal to 0,
           except for those in decomposition level j. Then, fill only decomposition
           level j with its maximum coefficient increased by 200% and spaced
           according to the scale;
        ii. alternating signs scenario: make all wavelet coefficients equal to 0,
           except for those in decomposition level j. Then, fill only decomposition
           level j with its maximum coefficient increased by 200% and spaced
           according to the scale, alternating the signs between positive and negative;
    end
end
```

In the pseudo-code above, the maximum number of decomposition levels is set to 8 to limit the number of simulations while keeping the most interesting scales, given the

comparison to limits established in EN 13848-5, which covers wavelengths ranging from 3 m to 70 m. A total of $26 \times 8 \times 2 = 416$ simulations were performed. The reconstruction is made for all the four irregularity signals, considering one mother wavelet from the list and one decomposition level of that wavelet at a time. The coefficients in all the decomposition levels are set to zero at first. Then, for the level being reconstructed, the reconstruction is based on selecting the highest wavelet coefficient observed for that decomposition level increased by 200%. The wavelet coefficient is then replicated 32 times, which corresponds to 64 m intervals when the track length of 2048 m is taken into consideration. The 2048 m track corresponds to $8192 = 2^{13}$ observations (or coefficients), with spacing between observations of 0.25 m. This means, for example, that for the detail coefficients in the first decomposition level, containing $4096 = 2^{12}$ coefficients (related to changes on a scale of 0.5 m), the value of the highest coefficient will be replicated every 128^{th} entry in the entries corresponding to that decomposition level in the vector of wavelet coefficients. Similarly, for the second decomposition level, containing $2048 = 2^{11}$ coefficients (related to changes on a scale of 1 m), the value of the highest coefficient will be replicated every 64^{th} entry in the entries corresponding to that decomposition level in the vector of wavelet coefficients. For all the other decomposition levels, up to level 8, the reconstruction follows the same procedure. In particular, for level 8, the $32 = 2^5$ coefficients are related to changes on a scale of 64 m), resulting in the most stretched wavelet. Lastly, two different scenarios for each wavelet and each decomposition level were simulated, where scenario (i.) keeps the original sign of the coefficient and scenario (ii.) alternates between positive and negative signs.

Table 5.2 brings the number of coefficients found at each decomposition level (from high-frequency detail coefficients 'd1' to low-frequency detail coefficients 'd8') plus the scale coefficient 's8', the window length in meters covered by each of the coefficients in each decomposition level, and the number of entries that will be set to zero between each nonzero coefficient at each decomposition level (column "Replicated at every x entries").

It should be noted that the division of the total number of coefficients in a particular decomposition level by the number of entries in the column "Repeated at every x^{th} entry" always results in 32, so that each simulation will be based on a signal reconstructed with only 32 coefficients, coming from a particular wavelet and a pre-defined decomposition level.

Table 5.2: Reconstruction scheme for each decomposition level

Decomposition Level	No. of coefficients	Window Length (m)	Repeated at every x^{th} entry
d1	4096	0.5	128
d2	2048	1	64
d3	1024	2	32
d4	512	4	16
d5	256	8	8
d6	128	16	4
d7	64	32	2
d8	32	64	1
s8	32	64	1

5.4.5 Mother Wavelet Selection

The first step of WA is to choose a mother wavelet. The DWT coefficients represent the energy level of the decomposed signal in each frequency band, with the first iteration having the high-frequency coefficients or fine-detailed features, moving to iterations linked to low frequency or coarse features. A wavelet transform that uses a few coefficients with a large percentage of total signal energy is desired, which means that it can compress information without losing signal energy. This implies that when the signal is reconstructed, many of the small coefficients can be omitted without losing significant information [125]. Scale is also an important aspect. A major frequency component occurring at a particular location indicates that the mother wavelet at that scale is similar or close to the signal in that location. As a result, the DWT has a large value at that location and scale [142].

Under the current experimentation scheme, different wavelet families were tested: haar, daubechies (d), symlet (s) and coiflet (c), in a total of 26 wavelets. For simplicity, they are further denominated by using the first letter of the wavelet family name (i.e., d, s or c) followed by the number of vanishing moments, with the exception of the haar wavelet

(also known as d1), as follows: haar, d2, d4, d6, d8, d10, d12, d14, d16, d18, d20, s2, s4, s6, s8, s10, s12, s14, s16, s18, s20, c6, c12, c18, c24, c30.

All these wavelets are orthogonal, and although they cannot be described by analytical expressions, they can be fairly described via filter coefficients. The exception is again, the haar wavelet, which has an analytical expression and can also be described via filters. All the other orthogonal wavelets are not symmetric. Orthogonality is a particularly interesting feature when dealing with non-stationarity signals, as the wavelet decomposition provides information on the variability of the signal's amplitude at the various scales with time/location. The number of vanishing moments also affects the support of the wavelet, which can be decisive in terms of feature detection: wavelets with smaller support will detect better closely spaced features. A wavelet with N vanishing moments is orthogonal to polynomials of degree $N - 1$. Vanishing moments are also intrinsically related to regularity: if a wavelet has N continuous derivatives, it must have at least $N + 1$ vanishing moments. Regularity relates to how many continuous derivatives a function has. Therefore, more vanishing moments imply more smoothness.

Minimum Entropy Criterion

A candidate mother wavelet was initially chosen based on the Shannon entropy criterion (Equation 5.9) further normalized by the coefficient values. A good mother wavelet candidate according to this criterion is the one which, with the least number of coefficients, has the most percentage of energy associated.

Figure 5.3 shows the entropy criterion for the 26 candidate wavelets and all the four irregularities signals.

According to this criterion, the following wavelets should be chosen: 'd16' for longitudinal level, 's18' for lateral alignment (although 'd14' had the same performance) and 's6' for both cross-level and gauge (although 'd6' had the same performance).

Figure 5.4 depicts, for each irregularity and mother wavelet selected according to the

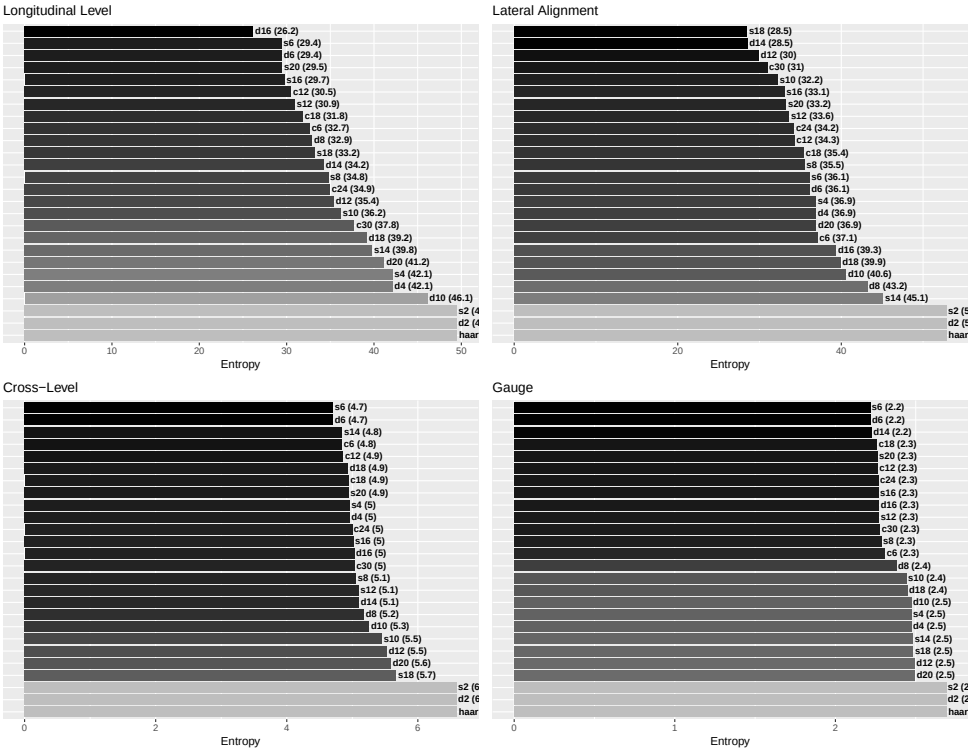

Figure 5.3: Wavelets entropy for each irregularity signal

minimum entropy criterion, the cumulative energy percentages added by each of the 5% highest energy coefficients (in a total of 410 coefficients), showing the level to which each coefficient belongs, i.e., if it is a detail (plus corresponding level) coefficient or a scaling coefficient.

From the depiction in Figure 5.4, it is clear that the longitudinal level and the lateral alignment signals behave differently than the cross-level and gauge signals. In particular, the coefficients with the highest energy associated with the first two come mostly from the detail levels 5, 6, 7 and 8, and the cumulative percentual increases slowly. For the cross-level and gauge signals, the highest coefficients in terms of energy are the scaling coefficients, and only a few coefficients will account almost entirely for the signal's total energy. In general, this analysis leads to the conclusion that the finer scales $1:4$ are not so important in terms of the signal's energy, hence, it is expected that these scales will not be relevant for fault detection.

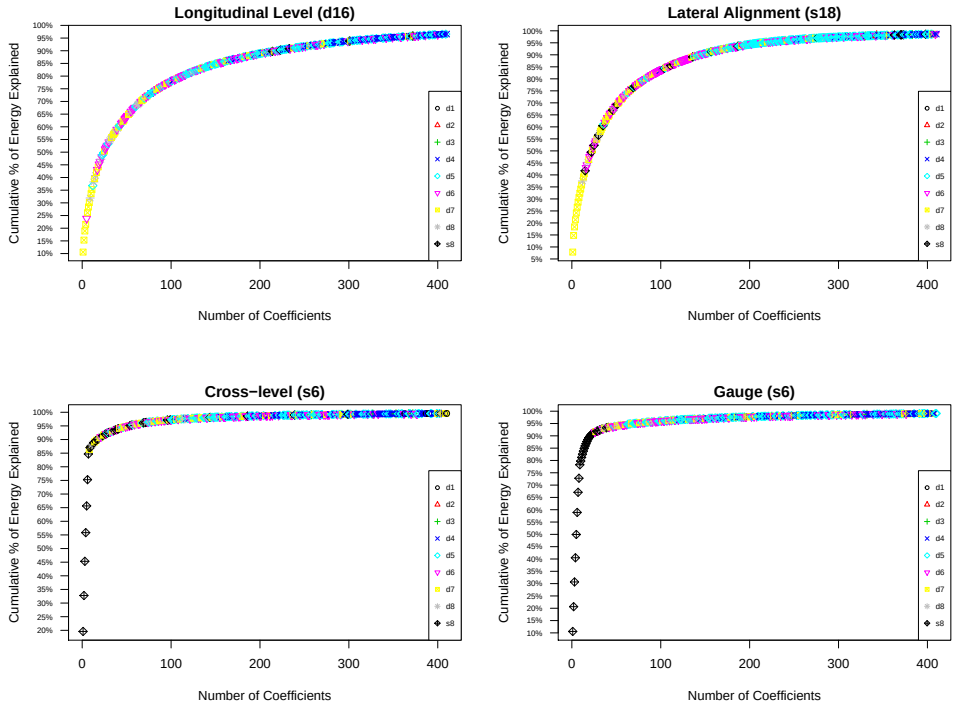

Figure 5.4: Depiction of minimum entropy mother wavelet coefficients

5.4.6 Vehicle Dynamics Simulations

The vehicle dynamics simulation package used is a commercial program developed specifically to study the railway dynamics, Vampire®. It requires as inputs the track irregularities and track layout, as well as vehicle inputs. One observation here is that the software uses a single signal for the alignment and longitudinal irregularities, and thus the average between the values in the two rails is supplied.

The vehicle dynamics simulations output comes in the form of a maximum of Y/Q per 1 m segment of the simulated track, i.e., the maximum value for the ratio Y/Q per 1 m of track segment. Figure 5.5 displays the results in terms of a maximum of Y/Q in the y-axis and the distance 109.3575 km to 111.4055 km in the x-axis obtained for the original track measurements. As it should be evident from Figure 5.5, according to the limit established in Equation 5.10 and highlighted in Figure 5.5 by the red dotted line, the measured irregularities do not result in dangerous Y/Q levels, as all values obtained fall far

from the red dotted line.

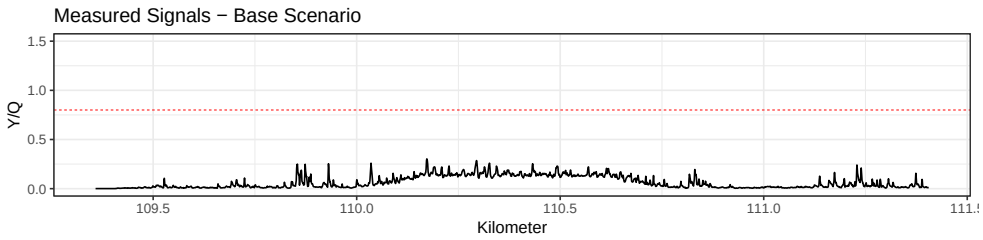

Figure 5.5: Vampire Y/Q output computed from real measured signals

Most "Penalizing" Wavelet Criterion

In addition to the wavelet obtained by the minimum entropy criterion, another experiment to find the most "penalizing" wavelet was simulated. The most "penalizing" refers to a wavelet that can be associated with the highest response of Y/Q under some conditions. The experiment assumes that, starting with level 1 of the finest details, if a signal is reconstructed based only on each decomposition level frequency band, different signals in different frequencies and their effects in the Y/Q can be further isolated and investigated.

A total of $26 \times 8 \times 2 = 416$ simulations were performed according to Algorithm 1. Table B.1 in Appendix B shows the simulations results in terms of the number of times the Y/Q was greater than the limit of 0.8 for each wavelet, decomposition level and scenario simulated. Table B.2 shows the simulations results in terms of the mean Y/Q observed. After analyzing the results, the most penalizing wavelet in terms of Y/Q was 'd4', although 's4' performed similarly. Figure 5.6 presents the results for the $8 \times 2 = 16$ simulations associated with the wavelet 'd4', where the plots on the left side correspond to scenario 'i. keep the original sign of the coefficient' and the plots on the right side correspond to scenario 'ii. alternate between positive and negative signs'. Each row corresponds to a different decomposition level, starting with the finest details (level 1) at the top, down to the level with coarser details (level 8) at the bottom.

For 'd4' (Figure 5.6) the Y/Q limit of 0.8 established in Equation 5.10 is surpassed 219

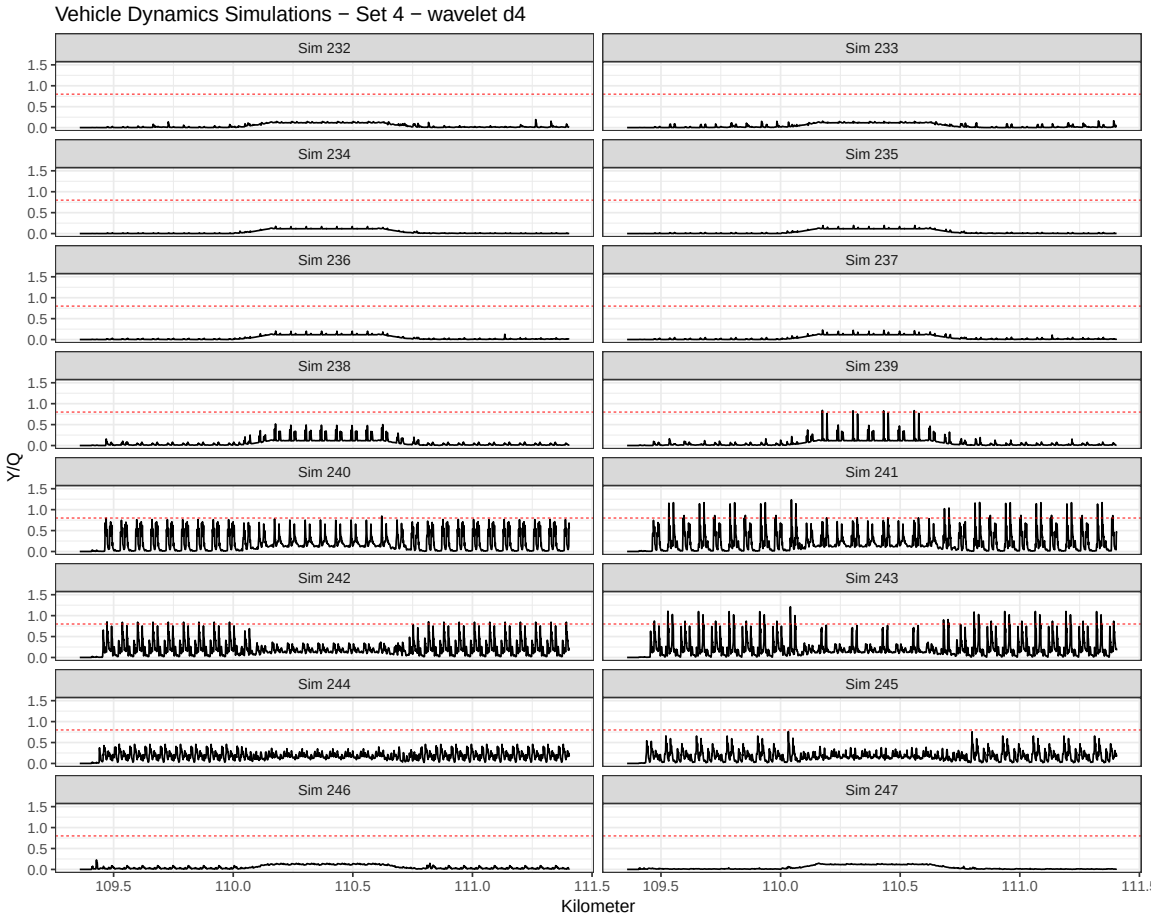

Figure 5.6: Most ”penalizing” wavelet ’d4’ simulations

times, 39 in decomposition levels 5 and 6 for scenario i and 180 times in decomposition levels 4, 5 and 6 for scenario ii. This wavelet was the one associated with the highest mean value for Y/Q, where 0.111 was the overall mean, 0.110 was the mean for scenario i and 0.113 was the mean for scenario ii. For comparison purposes, the minimum entropy wavelet for the longitudinal level, ’d16’ is presented in Figure 5.7.

For ’d16’ (Figure 5.7) the limit established in Equation 5.10 is never greater than 0.8 (see Table B.1) and the average responses are clearly much smaller than those observed for ’d4’: 0.078 was the value found for the overall mean and the mean of scenarios i and ii (see Table B.2).

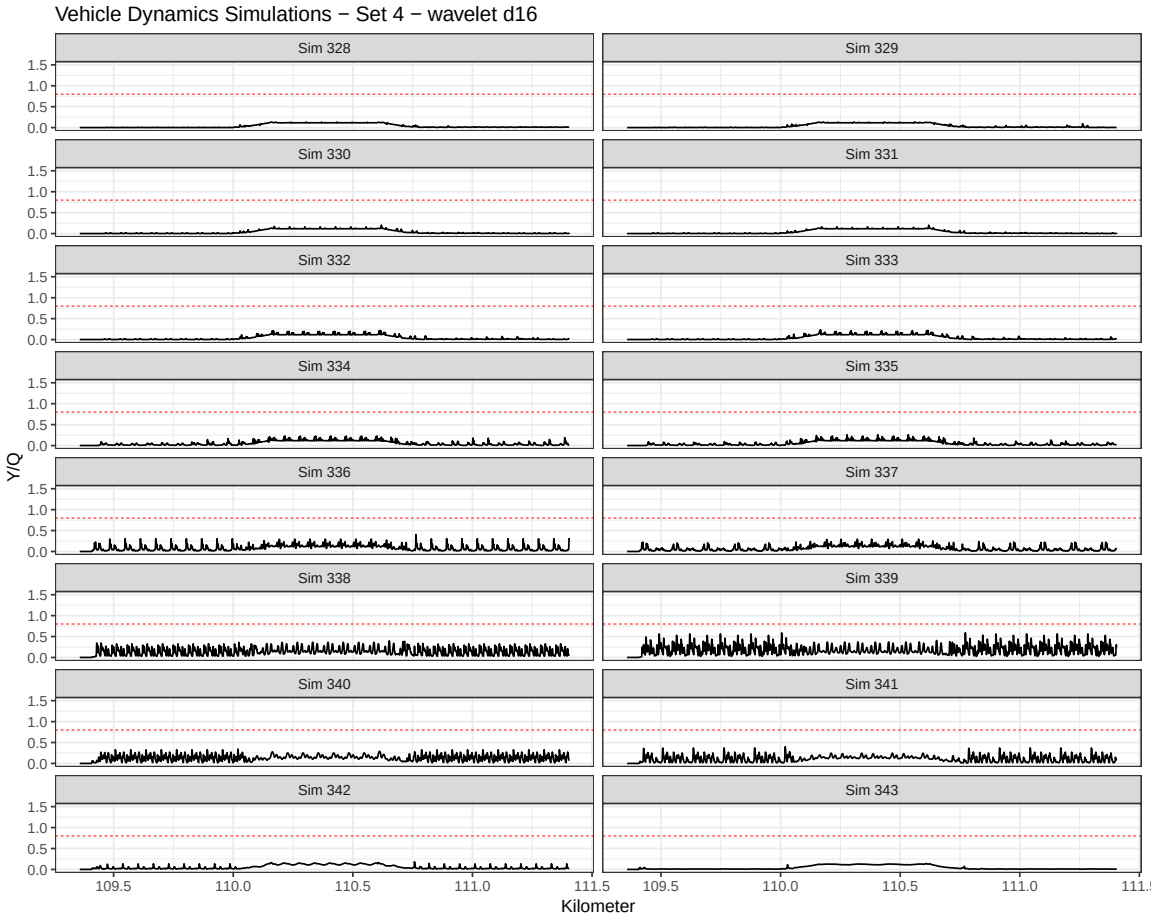

Figure 5.7: Minimum entropy wavelet 'd16' simulations

Phase 1 Simulation Results

For phase 1 simulations, the signals for all irregularities were reconstructed based on 50%, 100% and 200% increase in all wavelet coefficients, resulting in 3 simulations showing the response in terms of Y/Q for each wavelet tested. The simulation results for the irregularity signals reconstructed with the minimum entropy wavelets (i.e. 'd16' for longitudinal level, 's18' for lateral alignment and 's8' for cross-level and gauge) are shown on the left side of Figure 5.8 and the results obtained using the most "penalizing" wavelet 'd4' are shown on the right side.

From Figure 5.8 it should be clear that the results are almost identical. Indeed, Table 5.3 shows the results, for each scenario, in terms of the mean, the maximum and the variance

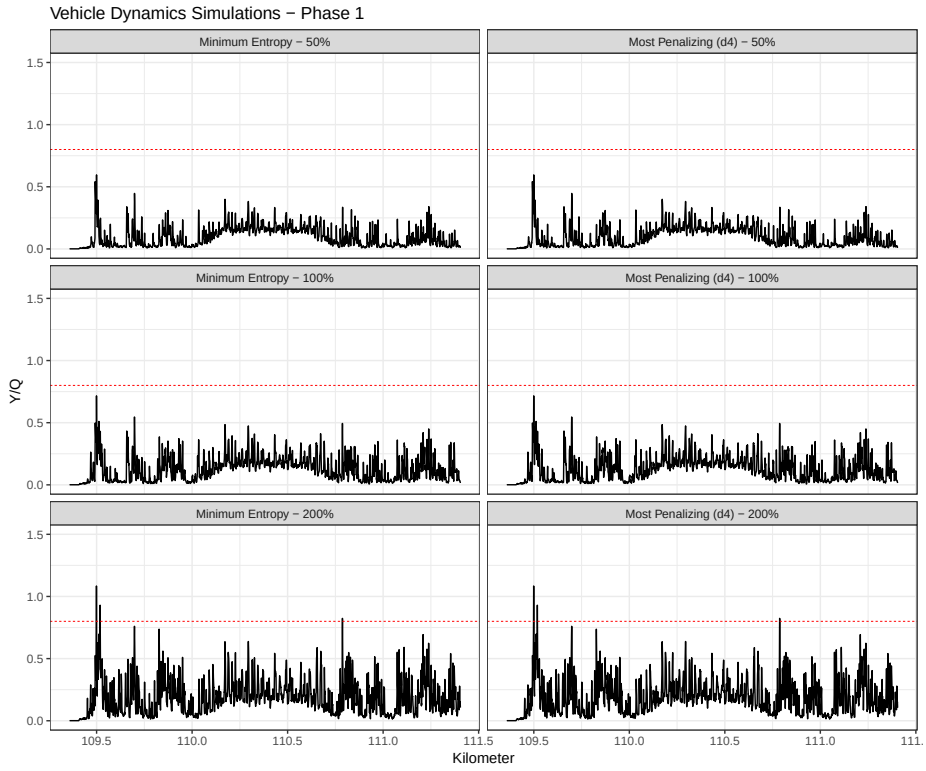

Figure 5.8: Phase 1 - Minimum entropy and most penalizing wavelets simulation results

of the Y/Q observed as well as the number of times the Y/Q was greater than the limit of 0.8. As can be noted, up to two decimal places, the results are the same.

Table 5.3: Results for phase 1 simulations

Scenario	Mean Y/Q	Max Y/Q	Variance Y/Q	number of $Y/Q > 0.8$
Minimum Entropy - 50%	0.10	0.60	0.01	0
Minimum Entropy - 100%	0.14	0.72	0.01	0
Minimum Entropy - 200%	0.20	1.08	0.02	6
Most Penalizing (d4) - 50%	0.10	0.60	0.01	0
Most Penalizing (d4) - 100%	0.14	0.72	0.01	0
Most Penalizing (d4) - 200%	0.20	1.08	0.02	6

The results obtained for phase 1 show that reconstructing the signals by increasing all coefficients in all decomposition levels by the same amount/percentage does not create different responses in terms of Y/Q for different wavelets. Although increasing coefficients by 200% of their measured values leads to peaks in the Y/Q extrapolating the limit of 0.8 established according to Equation 5.10, the results in phase 1 are inconclusive from the perspective of indicating which irregularity(ies) and/or scale(s) contributed the most in the

observed results. Hence, the investigation proceeds to phase 2.

Phase 2 Simulation Results

In phase 2, for each one of the 13 decomposition levels, reconstruction of all four irregularities signals were made one decomposition level at a time, while keeping the other scales untouched, based on 50%, 100%, and 200% increase in the wavelet coefficients. This resulted in $3 \times 13 = 39$ simulations for each wavelet tested.

Table 5.4 shows the results, for each wavelet, decomposition level, and scenario, in terms of the mean, the maximum, and the variance of the Y/Q observed as well as the number of times the Y/Q was greater than the limit of 0.8.

Table 5.4: Results for phase 2 simulations

Scenario	50% increase				100% increase				200% increase			
	Mean Y/Q	Max Y/Q	Variance Y/Q	Number of $Y/Q > 0.8$	Mean Y/Q	Max Y/Q	Variance Y/Q	Number of $Y/Q > 0.8$	Mean Y/Q	Max Y/Q	Variance Y/Q	Number of $Y/Q > 0.8$
Minimum Entropy - lvl 1	0.07	0.30	0.00	0	0.07	0.30	0.00	0	0.07	0.30	0.00	0
Most Penalizing (d4) - lvl 1	0.07	0.30	0.00	0	0.07	0.30	0.00	0	0.07	0.30	0.00	0
Minimum Entropy - lvl 2	0.07	0.30	0.00	0	0.07	0.30	0.00	0	0.07	0.30	0.00	0
Most Penalizing (d4) - lvl 2	0.07	0.30	0.00	0	0.07	0.31	0.00	0	0.07	0.32	0.00	0
Minimum Entropy - lvl 3	0.07	0.31	0.00	0	0.07	0.31	0.00	0	0.07	0.32	0.00	0
Most Penalizing (d4) - lvl 3	0.07	0.31	0.00	0	0.07	0.32	0.00	0	0.08	0.36	0.00	0
Minimum Entropy - lvl 4	0.07	0.34	0.00	0	0.08	0.37	0.00	0	0.10	0.65	0.01	0
Most Penalizing (d4) - lvl 4	0.07	0.34	0.00	0	0.08	0.39	0.01	0	0.10	0.62	0.01	0
Minimum Entropy - lvl 5	0.07	0.33	0.00	0	0.09	0.39	0.01	0	0.11	0.55	0.01	0
Most Penalizing (d4) - lvl 5	0.08	0.40	0.00	0	0.10	0.59	0.01	0	0.14	0.96	0.02	7
Minimum Entropy - lvl 6	0.08	0.39	0.00	0	0.09	0.44	0.01	0	0.14	0.72	0.01	0
Most Penalizing (d4) - lvl 6	0.08	0.48	0.01	0	0.10	0.44	0.01	0	0.14	0.59	0.01	0
Minimum Entropy - lvl 7	0.08	0.46	0.00	0	0.09	0.43	0.01	0	0.11	0.47	0.01	0
Most Penalizing (d4) - lvl 7	0.08	0.35	0.00	0	0.09	0.45	0.01	0	0.12	0.50	0.01	0
Minimum Entropy - lvl 8	0.07	0.41	0.00	0	0.07	0.46	0.00	0	0.08	0.43	0.00	0
Most Penalizing (d4) - lvl 8	0.07	0.31	0.00	0	0.07	0.35	0.00	0	0.07	0.44	0.00	0
Minimum Entropy - lvl 9	0.07	0.30	0.00	0	0.07	0.44	0.00	0	0.07	0.42	0.00	0
Most Penalizing (d4) - lvl 9	0.07	0.30	0.00	0	0.07	0.30	0.00	0	0.07	0.30	0.00	0
Minimum Entropy - lvl 10	0.07	0.31	0.00	0	0.07	0.31	0.00	0	0.07	0.30	0.00	0
Most Penalizing (d4) - lvl 10	0.07	0.30	0.00	0	0.07	0.30	0.00	0	0.07	0.30	0.00	0
Minimum Entropy - lvl 11	0.07	0.31	0.00	0	0.07	0.31	0.00	0	0.07	0.31	0.00	0
Most Penalizing (d4) - lvl 11	0.07	0.30	0.00	0	0.07	0.30	0.00	0	0.07	0.29	0.00	0
Minimum Entropy - lvl 12	0.07	0.46	0.00	0	0.07	0.46	0.00	0	0.07	0.36	0.00	0
Most Penalizing (d4) - lvl 12	0.07	0.31	0.00	0	0.07	0.31	0.00	0	0.07	0.31	0.00	0
Minimum Entropy - lvl 13	0.07	0.32	0.00	0	0.08	0.43	0.00	0	0.08	0.38	0.00	0
Most Penalizing (d4) - lvl 13	0.07	0.30	0.00	0	0.07	0.48	0.00	0	0.07	0.43	0.00	0

A quick inspection of Table 5.4 reveals that the only simulation that generated nonzero output for the column "Number of $Y/Q > 0.8$" was the one associated with the most "penalizing" wavelet 'd4' in the decomposition level 5 and 200% increase scenario. This was also the simulation which had the largest Y/Q mean and the largest variance. This is an important result as it shows that decomposition level 5 along with wavelet 'd4' was the only combination that generated peaks of Y/Q, even though the other decomposition levels and even another wavelet were subject to a similar experimental setup.

Since the simulation results imply in 3 panels for each decomposition level and each wavelet, and there are 13 decomposition levels, Figure 5.9 only shows a visual comparison

of the results obtained for level 5, where the limit value for Y/Q surpasses the limit in Equation 5.10 seven times around 109.5 km for the 200% increase scenario. On the left side of Figure 5.9 are located the simulation results for the irregularity signals reconstructed with the minimum entropy wavelets (i.e. 'd16' for longitudinal level, 's18' for lateral alignment and 's8' for cross-level and gauge), and on the right side are located the results obtained using the most "penalizing" wavelet 'd4'.

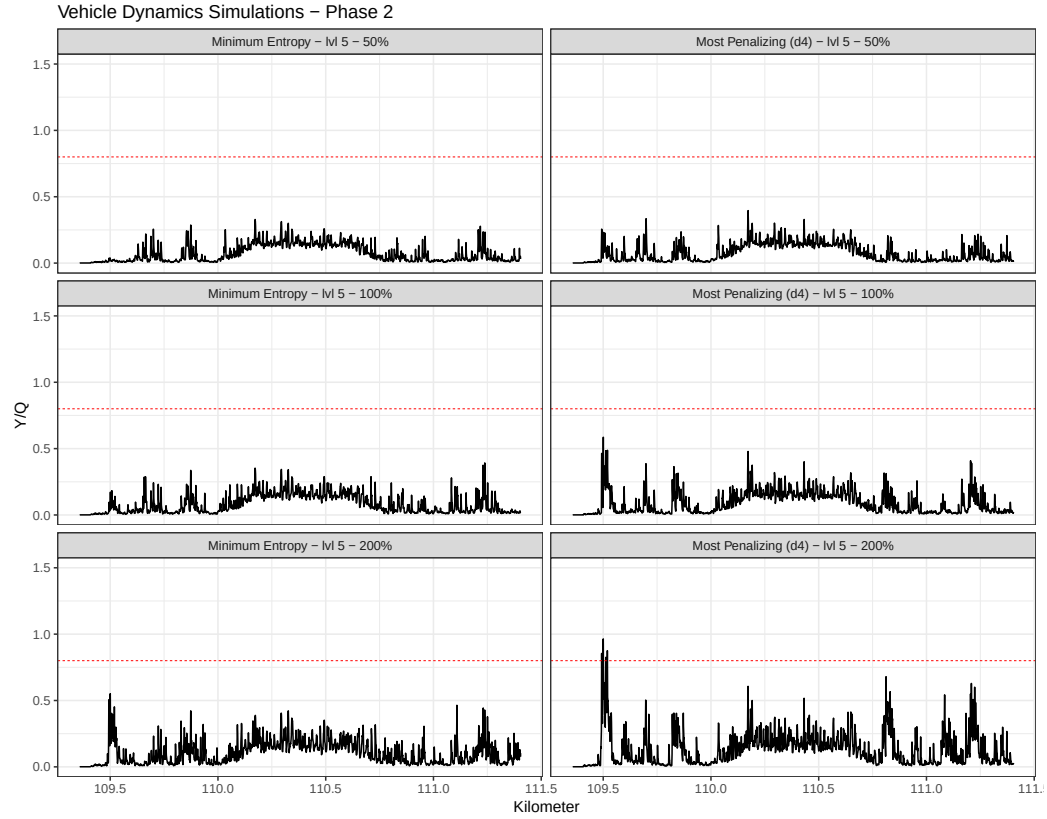

Figure 5.9: Phase 2 - Minimum entropy and most penalizing wavelets simulation results for level 5

The results from phase 2 put into evidence that different wavelength content existing in different decomposition levels may excite masses in the train and in the track in different manners, and in some specific cases, this may lead to an increased Y/Q response. But, most importantly, the fact the different wavelets subject to the same experimental setup generated different results when the excitation occurred in some specific decomposition levels reveals that it is worth investigating whether the "shape" intrinsic to the wavelet is

playing an important role in the results observed. This is the goal of phase 3, which will be explored next.

Phase 3 Simulation Results

After observing and analyzing the results for phases 1 and 2, phase 3 proceeds with the investigation of the most "penalizing" wavelet 'd4'. The experimental setup is as follows: for each one of the 13 decomposition levels, reconstruction is made for only one irregularity signal and one decomposition level at a time, while keeping the other irregularities and scales untouched, based on 50%, 100% and 200% increase in the wavelet coefficients. This results in $3 \times 13 \times 4 = 156$ simulations for the 'd4' wavelet.

All the simulations for the phase 3 - 'd4' wavelet were plotted in Appendix C, where Figure C.1 refers to decomposition level 1, Figure C.2 refers to decomposition level 2, Figure C.3 refers to decomposition level 3, Figure C.4 refers to decomposition level 4, Figure C.5 refers to decomposition level 5, Figure C.6 refers to decomposition level 6, Figure C.7 refers to decomposition level 7, Figure C.8 refers to decomposition level 8, Figure C.9 refers to decomposition level 9, Figure C.10 refers to decomposition level 10, Figure C.11 refers to decomposition level 11, Figure C.12 refers to decomposition level 12, and Figure C.13 refers to decomposition level 13. The most interesting results are again obtained for decomposition level 5, which will be further commented.

Figure 5.10 shows the phase 3 simulation results for decomposition level 5 under the most "penalizing" wavelet 'd4'.

In Figure 5.10, the first row of panels corresponds to the scenario where the coefficients were increased by 50% of their original values. The second row corresponds to an increase of 100% and the third row corresponds to an increase of 200%. Column 1 brings the simulations where only the longitudinal level signal at decomposition level 5 coefficients were changed according to the percentages on each row. For example, the first panel ('Sim 56') on the top left corresponds to the simulation output for the measured input

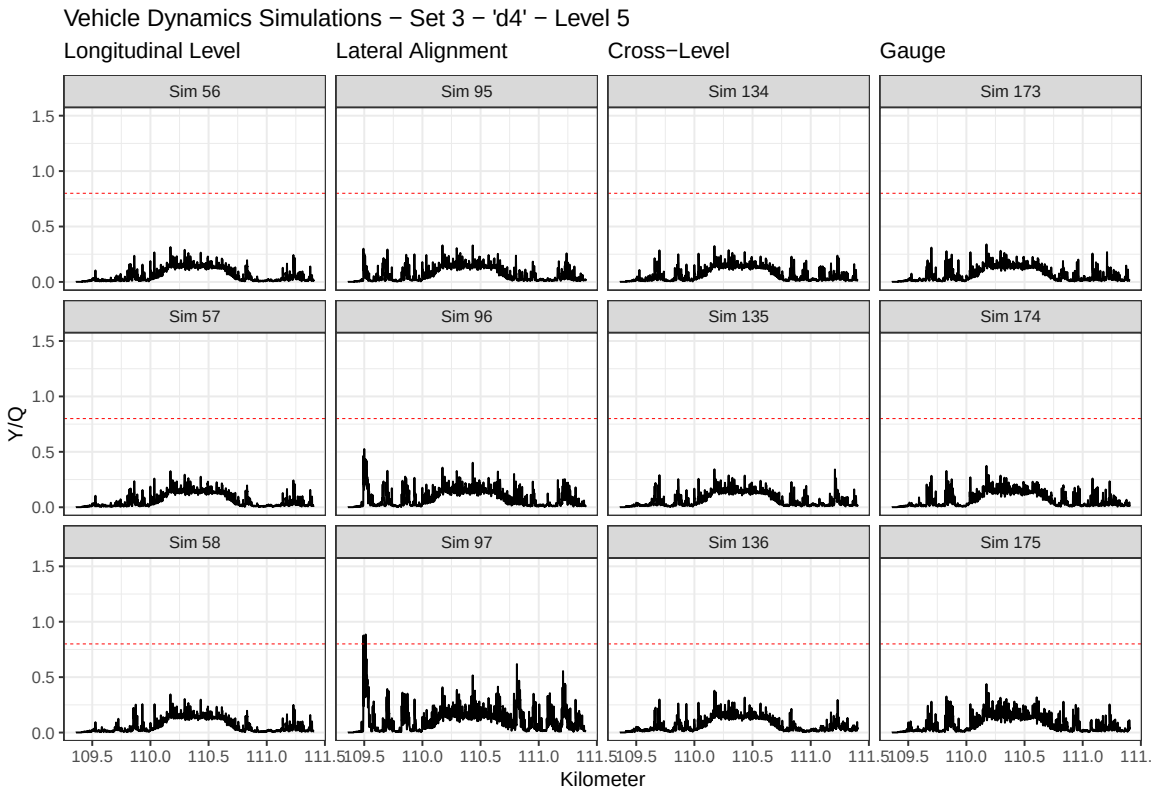

Figure 5.10: Phase 3 simulations - wavelet 'd4' - Level 5

signals of lateral alignment, cross-level and gauge, although it modifies the longitudinal level input by reconstructing the signal based on WT, increasing only the coefficients at decomposition level 5 by 50% of their original values (all other decomposition levels are kept the same). Column 2 assumes the same for the lateral alignment signals (changes only to lateral alignment at level 5, while keeping other signals and levels unchanged). Column 3 does the same for the cross-level and column 4 for the gauge.

The inputs which differ from the measured signals (in Figure 5.2) for each simulation, and considering the modifications applied only to decomposition level 5, are shown in Figure 5.11.

The reason why decomposition level 5 was chosen to be displayed is that, as suggested by a quick inspection of the plots in Appendix C, the output at level 5 and, more specifically, in the lateral alignment 200% simulated scenario, was the only one to present a response Y/Q exceeding the limit of 0.8. This means that a change in a single irregularity signal at a

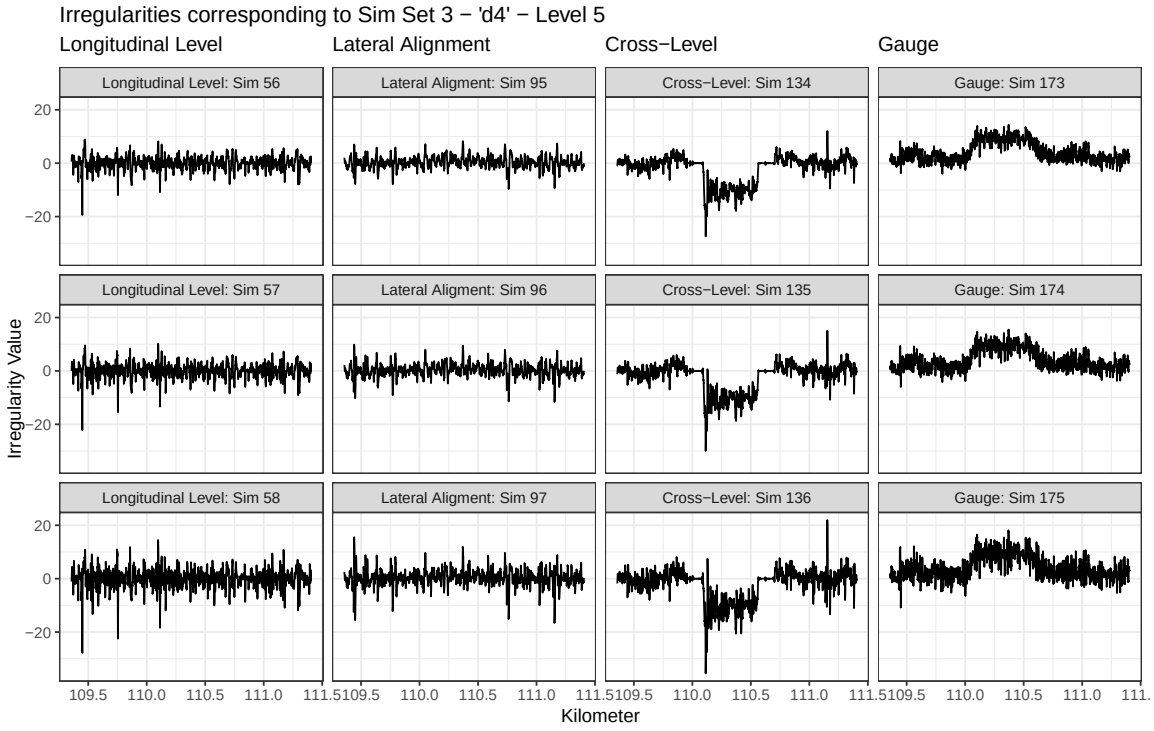

Figure 5.11: Input signals for phase 3 simulations - wavelet 'd4' - Level 5

single decomposition level was able to make a significant impact on the response, making it exceed the limit of the safety parameter. By using WA, it was possible to obtain the location of this defect in the lateral alignment signal, which makes it possible to conduct a localized maintenance action (e.g. localized tamping action). Interestingly enough, for the 120 km/h experimental speed, the IAL limit (Table 5.1) for the lateral alignment is not reached. Figure 5.12 shows the lateral alignment signal reconstructed with 'd4' by increasing level 5 coefficients by 50%, 100%, 200%, respectively, augmented at the maximum Y/Q output location for better visualization.

When the coefficients at level 5 got increased by 200% of their original values, the maximum amplitude obtained for the lateral alignment irregularity signal was 15.53 mm, which is still below the IAL of 17 mm established in EN 13848-5 (see Table 5.1). Similar behavior was not observed in other wavelets of the same family, nor other irregularity signals. In fact, phase 2 results showed that the analysis with the minimum entropy wavelets, for which the lateral alignment was reconstructed with 's18', did not generate

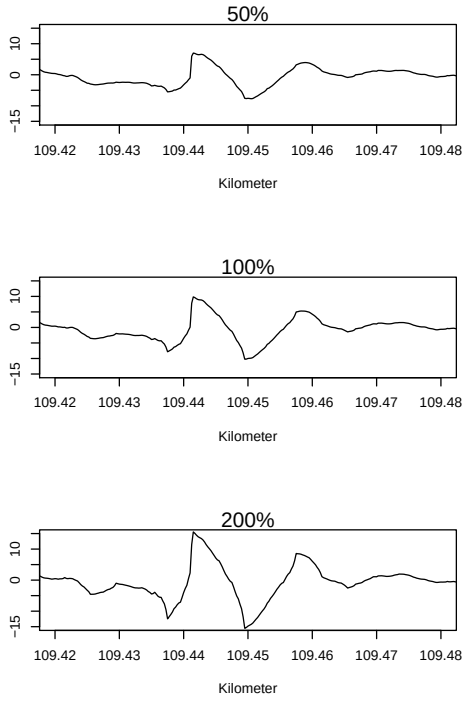

Figure 5.12: Lateral alignment signal reconstructed with 'd4' by increasing level 5 coefficients by 50%, 100%, 200%

any possible defects with $Y/Q > 0.8$. Even wavelets with a similar number of vanishing moments did not behave similarly. Figure 5.13 brings an example of phase 3 simulations reconstructed with 'd6', which has a close number of vanishing moments when compared to 'd4'.

One possible explanation, as previously suggested in phase 2 discussion, could be related to the similarity between the irregularity signal and the wavelet itself. Figure 5.14 brings the different formats or "shapes" for some of the wavelets investigated according to their decomposition level, for levels 1 through 8.

Visual comparison of the lateral alignment defect at the bottom of Figure 5.11 and the 'd4' filter at decomposition level 5 (first panel, row 5) in Figure 5.14 indeed reveals high similarity. The same shape is not observed for the other Daubechies in Figure 5.14, for which the reconstruction of the signals did not cause an increase of Y/Q above the limit of 0.8, even under the same experimental conditions. A simple visual inspection of the

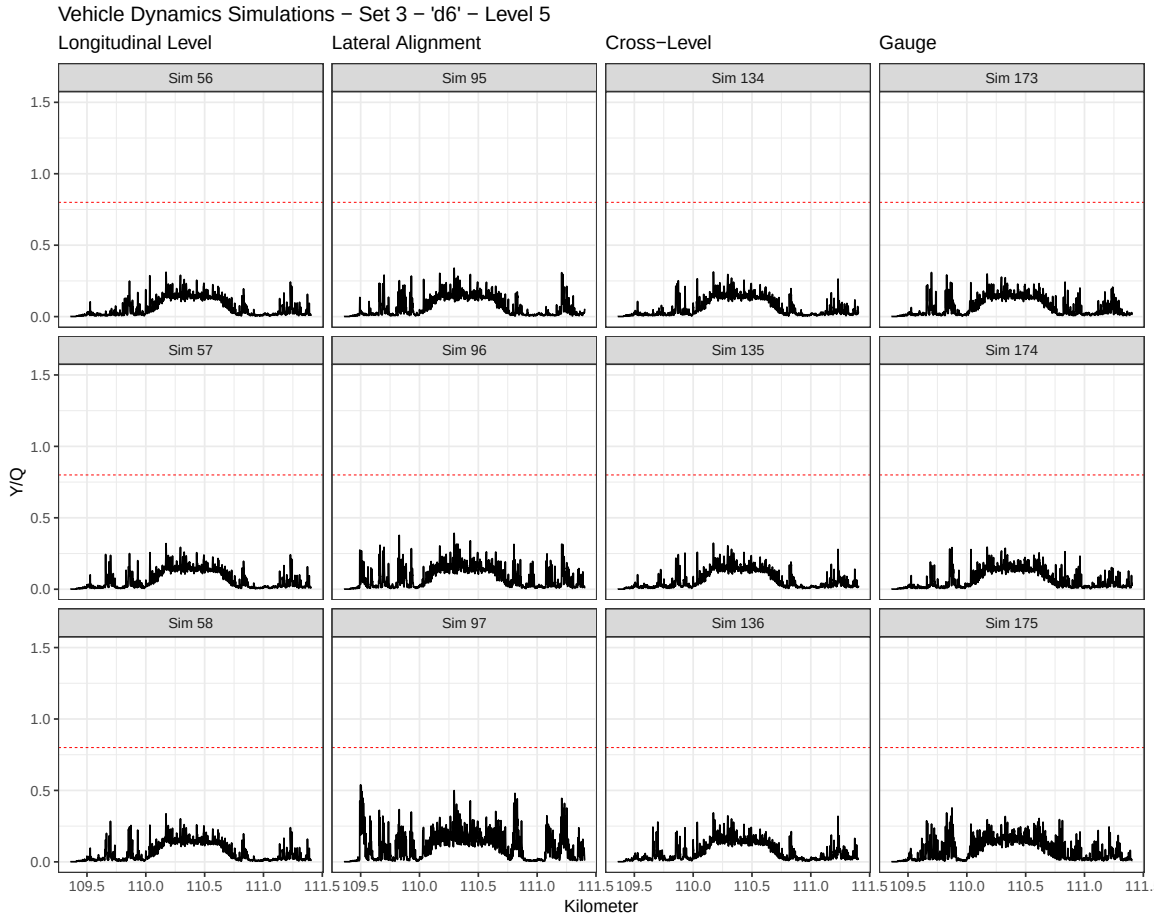

Figure 5.13: Input signals for phase 3 simulations - wavelet 'd6' - Level 5

figures in Appendix C suggests that the lateral alignment signal reconstructed with wavelet 'd4' promotes more excitation of the different train and track masses, which seems to be particularly harmful in levels $5-8$, but especially level 5. For the other irregularities, except the cross-level which does not have an IAL established, even though the last scenario with the 200% increase in the coefficients has generated signals with some values close to the IAL in the standard EN 13848-5 (Table 5.1), the same increase in the response was not observed and the Y/Q resulting signals were all still under control, with mean and variance values that did not vary significantly from one simulation to another.

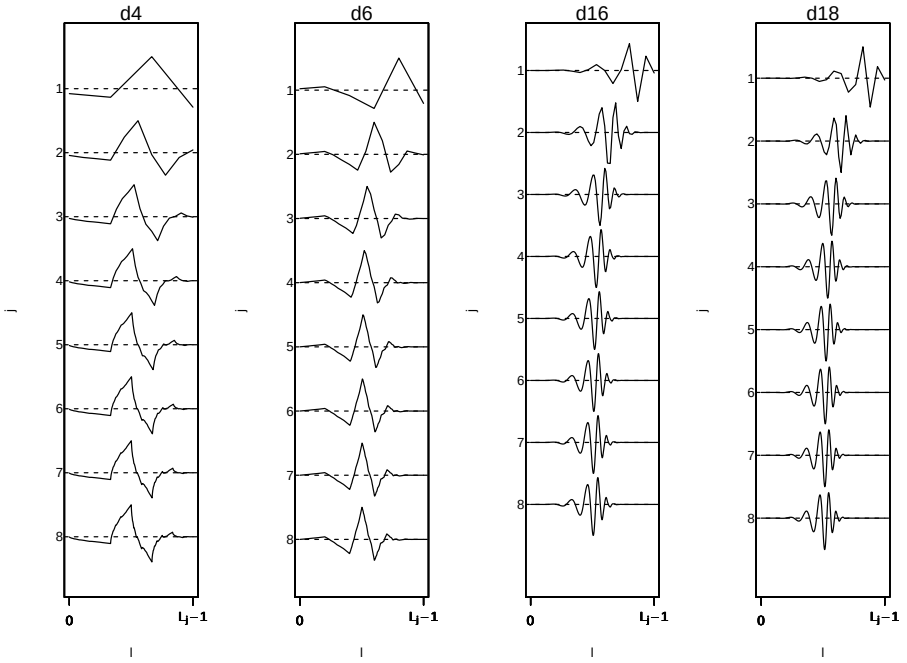

Figure 5.14: Filter coefficients for decomposition levels 1 through 8 of different wavelets

5.5 Conclusions and Further Research

This study explored the use of Wavelet Analysis in the statistical modeling of railway track irregularities. WA multiresolution capability makes it a powerful joint location-frequency analysis technique, particularly suitable for non-stationary or transient signals, which has resulted in WA gaining increasing attention in the research field of fault diagnostics. The investigation presented in this chapter is a contribution to a recent debate topic within track maintenance which refers to the use of predefined standard alert limits, such as those found in EN 13848-5 [1], and presents an alternative assessment method based on wavelets. Literature has already indicated that relying solely on point-wise comparisons with the current standard maintenance limits may be a flawed approach for various reasons, including, but not limited to: the averaging effect caused by considering long maintenance sections [9]; evidence of the existence of features in the track for which the vehicle responds to, although they lie within the current alert limits [119]; and the importance of considering

dynamic responses at the wheel-rail (or train-track) interface [88].

In particular, the assessment method proposed in this chapter combined WA and vehicle dynamics simulations to identify track faults by using the traditional Nadal's safety criterion for derailment, known as Y/Q. The goal was to find evidence of wavelets and respective scales which could be associated to increased Y/Q values, especially those surpassing the established limit of 0.8 for the safety of trains operations, even when the original track geometric signals measured at the extrapolation points had not surpassed the immediate action limit (IAL) suggested in the conventional track maintenance standards, in this case, the European Standard EN 13848-5 [1]. Based on the assumption that the various defects have a range of varying wavelength and frequency content, their observance in some specific "shapes" in the original signal could be an indicator of track faults that are potentially more penalizing in terms of Y/Q. Therefore, WA was a natural choice due to some interesting properties of wavelets: they are well-localized (in time/location and in scale/frequency), they preserve, but unbalance the energy in the data and they are versatile, coming in a wide range of shapes and formats. In addition, all wavelets used in this chapter were orthogonal.

The vehicle dynamics simulation experiment was based on three phases, upon the choice of wavelets to be tested according to two main criteria: minimum entropy and most "penalizing". Regarding minimum entropy wavelets, the chosen wavelets according to the criterion were: 'd16' for longitudinal level, 's18' for lateral alignment and 's6' for cross-level and gauge. With respect to the most "penalizing" criterion in terms of Y/Q, which was based on measuring the highest simulated response of Y/Q after experimentation with 26 orthogonal wavelets and reconstruction of signals at different decomposition levels, the chosen wavelet, for all signals, was 'd4'.

For phase 1 simulations, the signals for all irregularities were reconstructed based on 50%, 100% and 200% increase in all wavelet coefficients at all decomposition levels. The results, although suggestive that an increase in the amplitudes of all coefficients will lead

to increased Y/Q response, were inconclusive from the perspective of indicating which irregularity(ies) and/or scale(s) contributed the most in the observed results. Moreover, the reconstruction based on all coefficients of the wavelets tested resulted in identical reconstructed irregularity signals, which also did not allow to isolate the effect (if any) of the wavelet being tested.

In phase 2, for each one of the 13 decomposition levels, reconstruction of all four irregularities signals were made one decomposition level at a time, while keeping the other scales untouched, based again on 50%, 100% and 200% increase in the wavelet coefficients. Only one simulation generated response $Y/Q > 0.8$, and that result was associated with the most "penalizing" wavelet 'd4' in the decomposition level 5 and 200% increase scenario. This was also the simulation which had the largest Y/Q mean and the largest variance. This was an important result as it showed that decomposition level 5 along with the most "penalizing" wavelet 'd4' was the only combination that generated peaks of Y/Q above the limit, even though other decomposition levels and even other wavelets were subject to a similar experimental setup.

Phase 3 provided a more in-depth investigation of the results obtained in phase 2. The experimental setup for phase 3 was based on the reconstruction of a single irregularity signal and one decomposition level at a time, while keeping the other irregularities and scales unchanged (ceteris paribus), based on 50%, 100% and 200% increase in the wavelet coefficients. The results for the most 'penalizing' wavelet 'd4' agreed with from those of phase 2, and the output at decomposition level 5 and, more specifically, in the lateral alignment 200% simulated scenario, was the only one to present a response Y/Q exceeding the limit of 0.8. However, by isolating and reconstructing only one irregularity signal at a time, phase 3 revealed that the lateral alignment was the geometric signal that had the most significant impact on the observed Y/Q. Moreover, a comparison of 'd4' with another similar wavelet, 'd6', showed that the wavelet choice factor also played an important part, as the reconstructed scenarios with 'd6' were not able to impact the Y/Q in the same way,

and no points extrapolating the limit were found. More interesting was the finding related to the potential defect created in the lateral alignment signal by reconstructing decomposition level 5 with coefficients increased by 200%, resulting in the observed peak of 15.53 mm for lateral alignment around 109.5 km. In this case, for the 120 km/h experimental speed, the IAL limit from EN 13848-5 of 17 mm is not reached. One possible explanation could be related to the similarity between the irregularity signal and the wavelet itself, and visual comparison of the lateral alignment defect and the 'd4' filter at decomposition level 5 clearly showed high similarity.

Two main conclusions that can be immediately inferred from this experimental investigation are: defects occurring in some scales/frequencies are potentially more harmful from the standpoint of Y/Q than other defects in other scales/frequencies, and the similarity of some particular geometric defect with the "shape" of some wavelet may be used to identify points where to target maintenance actions. Although, as emphasized in EN 13463 [145], it is unlikely that a single peak of Y/Q would cause a derailment, as this usually requires having high amplitudes of Y/Q along a significant portion of the track, these findings raise questions on the appropriateness of applying the same track maintenance standards to all train operating companies and track irregularities based on point-wise comparisons with alert limits.

Many other interesting aspects arise from this investigation and are worth being further investigated. The first aspect is related to the response in the transition-curve-transition segment. It seems that the oscillations of the response in that segment are still approximately within the range of the amplitudes observed in the other two "straight" segments, although the pattern of the oscillations changes. The second aspect is related to the alternating vs nonalternating sign simulations. These two scenarios were considered in the simulation experiment to find the most "penalizing" wavelet, but they were not included in phases 1, 2, or 3. This was due to the lack of physical interpretation and actual feasibility to obtain these in a real-world scenario. In fact, alternating signs could indicate,

for instance, whether a lateral alignment is more to the left or to the right, or whether a longitudinal level deficiency is positive or negative. However, the actual feasibility of encountering something oscillating between negative and positive in real tracks is a hypothesis that must be further investigated. What the most "penalizing" wavelet analysis indicated is that, for all wavelets tested, higher Y/Q responses were found in the alternating signs scenarios, in comparison to the nonalternating (or original) ones.

Future research that includes a more in-depth statistical validation and generalization of the results observed from the simulations is also desired, although this may not be straightforward. There are many other wavelets and defects to be investigated, and many more factors acting on the track which may affect the results. A trivial next step would be to create a Design of Computer Experiments to allow for simulation of different combinations of irregularities signals at varying frequencies to help to answer the question: how can wavelet analysis improve how inspection data from railway track irregularities is analyzed and interpreted?

Appendices

APPENDIX A

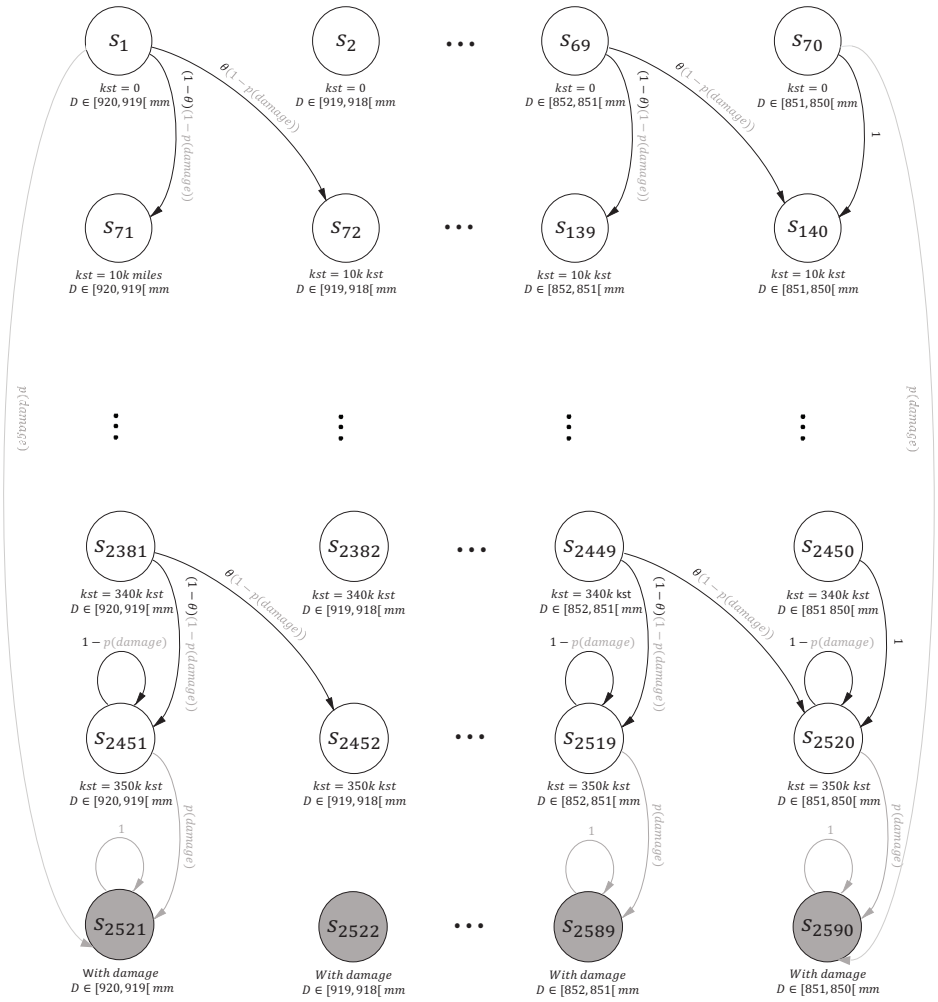

Figure A.1: State space division and transition probabilities for the 'Do nothing' action ($a = 1$)

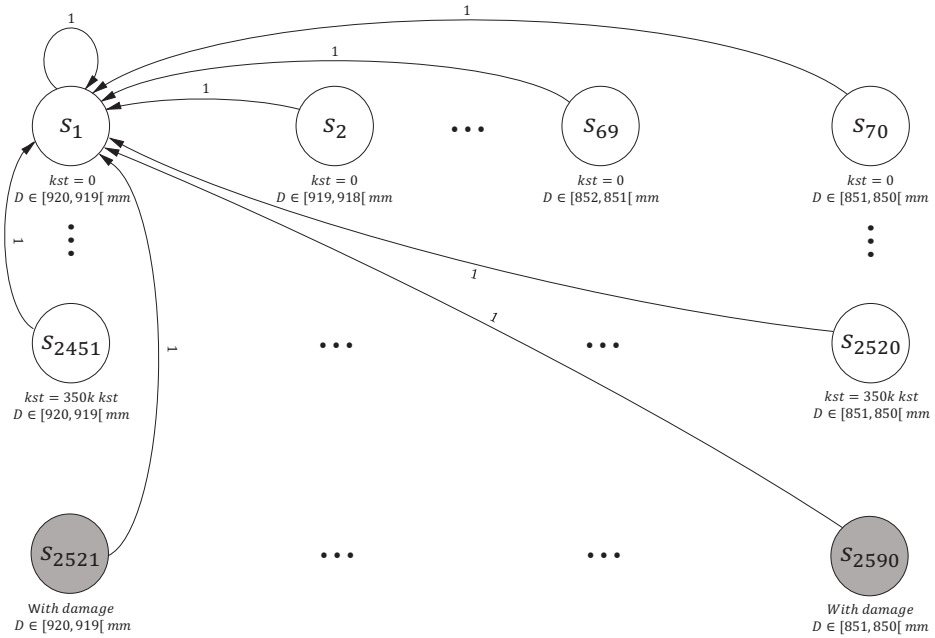

Figure A.2: State space division and transition probabilities for the 'Renewal' action ($a = 2$)

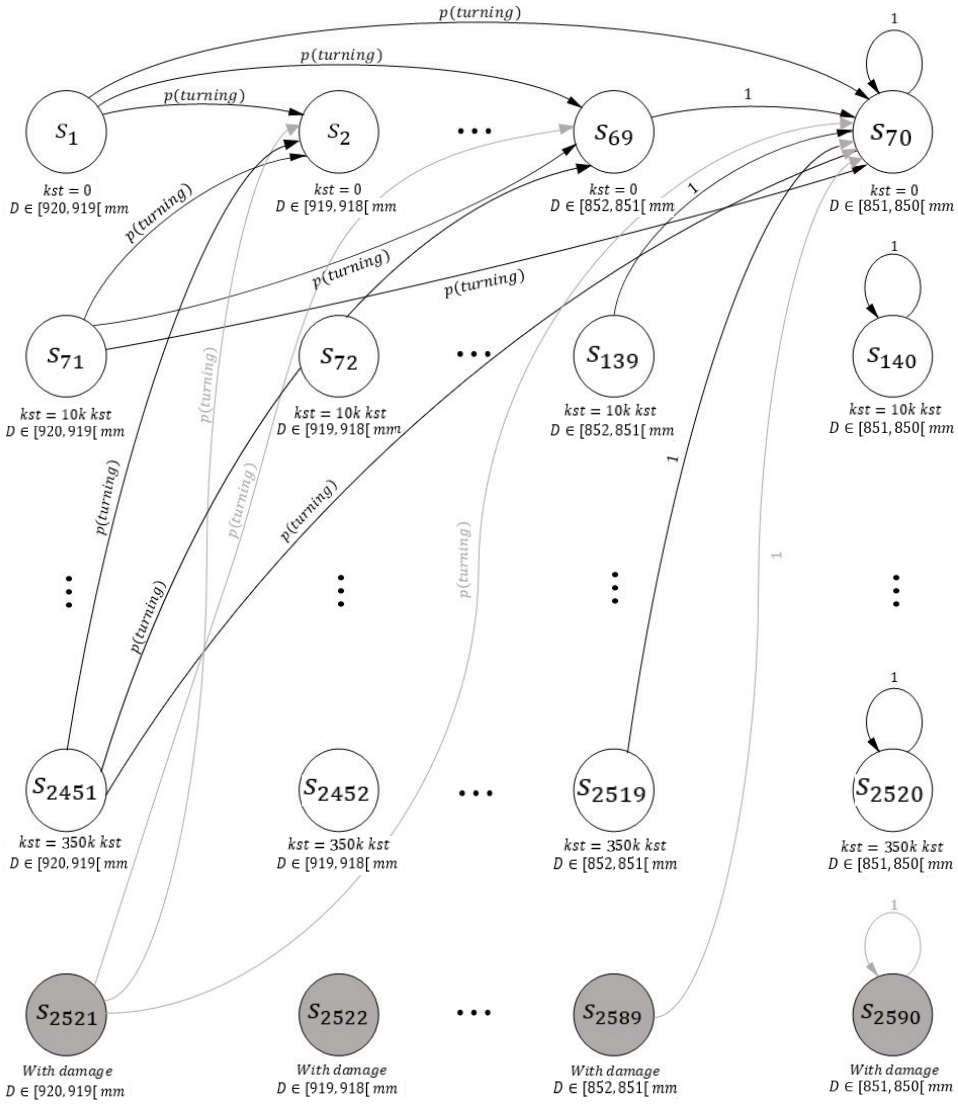

Figure A.3: State space division and transition probabilities for the 'Turning' action ($a = 3$)

APPENDIX B

Table B.1: Simulations results - number of times $Y/Q > 0.8$ for each wavelet, decomposition level and scenario simulated

	non-alternating signs - scenario i).									alternating signs - scenario ii).									
Wavelet/level	1	2	3	4	5	6	7	8	Subtotal	1	2	3	4	5	6	7	8	Subtotal	Total
d4	0	0	0	0	1	38	0	0	39	0	0	0	4	91	85	0	0	180	219
s4	0	0	0	0	1	38	0	0	39	0	0	0	4	91	85	0	0	180	219
c6	0	0	0	0	14	0	0	0	14	0	0	0	0	44	67	0	0	111	125
c12	0	0	0	0	18	0	0	0	18	0	0	0	0	22	0	0	0	22	40
s8	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	26	26
c18	0	0	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	22	22
d10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	18	18
haar	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	3	0	5	7
s2	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	3	0	5	7
d2	0	0	0	2	0	0	0	0	2	0	0	0	2	0	0	3	0	5	7
d14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
s20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
d6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table B.2: Simulations results - mean of Y/Q for each wavelet, decomposition level and scenario simulated

	non-alternating signs - scenario i).									alternating signs - scenario ii).									
Wavelet/level	1	2	3	4	5	6	7	8	Subtotal	1	2	3	4	5	6	7	8	Subtotal	Total
d4	0.046	0.039	0.042	0.065	0.214	0.221	0.19	0.057	0.109	0.048	0.04	0.043	0.066	0.235	0.241	0.183	0.044	0.113	0.111
s4	0.046	0.039	0.042	0.065	0.214	0.221	0.19	0.057	0.109	0.048	0.04	0.043	0.066	0.235	0.241	0.183	0.044	0.113	0.111
c6	0.043	0.058	0.043	0.059	0.189	0.24	0.188	0.058	0.11	0.045	0.06	0.045	0.06	0.203	0.261	0.184	0.046	0.113	0.111
s8	0.044	0.05	0.04	0.058	0.203	0.177	0.139	0.054	0.096	0.042	0.052	0.04	0.063	0.221	0.188	0.132	0.043	0.098	0.097
c12	0.042	0.054	0.047	0.057	0.206	0.16	0.149	0.054	0.096	0.04	0.054	0.05	0.058	0.227	0.171	0.133	0.043	0.097	0.097
s10	0.046	0.038	0.054	0.057	0.187	0.15	0.159	0.056	0.093	0.046	0.039	0.058	0.059	0.194	0.156	0.14	0.043	0.092	0.093
s2	0.066	0.05	0.049	0.081	0.138	0.087	0.185	0.113	0.096	0.062	0.047	0.049	0.067	0.129	0.124	0.154	0.059	0.086	0.091
d2	0.066	0.05	0.049	0.081	0.138	0.087	0.185	0.113	0.096	0.062	0.047	0.049	0.067	0.129	0.124	0.154	0.059	0.086	0.091
haar	0.066	0.05	0.049	0.081	0.138	0.087	0.185	0.113	0.096	0.062	0.047	0.049	0.067	0.129	0.124	0.154	0.059	0.086	0.091
s12	0.042	0.038	0.053	0.056	0.192	0.147	0.148	0.056	0.092	0.04	0.038	0.053	0.06	0.205	0.152	0.127	0.043	0.09	0.091
s16	0.042	0.045	0.044	0.054	0.189	0.131	0.152	0.056	0.089	0.04	0.049	0.046	0.057	0.206	0.132	0.129	0.043	0.088	0.088
c18	0.041	0.047	0.04	0.063	0.194	0.124	0.143	0.055	0.088	0.04	0.051	0.04	0.068	0.215	0.119	0.121	0.043	0.087	0.088
s20	0.042	0.039	0.041	0.051	0.19	0.128	0.153	0.056	0.087	0.041	0.038	0.041	0.053	0.204	0.132	0.124	0.043	0.084	0.086
s18	0.045	0.039	0.04	0.072	0.15	0.132	0.172	0.056	0.088	0.042	0.039	0.04	0.072	0.136	0.135	0.135	0.043	0.08	0.084
d10	0.038	0.04	0.041	0.065	0.067	0.201	0.121	0.084	0.082	0.038	0.039	0.041	0.065	0.074	0.259	0.111	0.052	0.085	0.083
d8	0.038	0.039	0.043	0.059	0.086	0.184	0.131	0.076	0.082	0.038	0.039	0.043	0.06	0.091	0.204	0.12	0.048	0.08	0.081
s14	0.039	0.039	0.041	0.055	0.096	0.19	0.116	0.079	0.082	0.038	0.038	0.041	0.055	0.1	0.211	0.101	0.049	0.079	0.081
d20	0.04	0.042	0.041	0.054	0.191	0.101	0.143	0.053	0.083	0.04	0.043	0.04	0.053	0.184	0.106	0.116	0.043	0.078	0.081
d12	0.038	0.038	0.042	0.059	0.068	0.142	0.171	0.076	0.079	0.04	0.038	0.042	0.058	0.074	0.205	0.147	0.047	0.081	0.08
d14	0.04	0.039	0.043	0.06	0.109	0.119	0.181	0.049	0.08	0.041	0.039	0.043	0.061	0.111	0.149	0.15	0.041	0.079	0.08
d16	0.04	0.04	0.045	0.058	0.09	0.153	0.144	0.058	0.078	0.041	0.04	0.045	0.057	0.08	0.2	0.12	0.042	0.078	0.078
c24	0.041	0.047	0.042	0.055	0.13	0.123	0.138	0.056	0.079	0.039	0.049	0.047	0.059	0.141	0.122	0.116	0.043	0.077	0.078
s6	0.039	0.038	0.046	0.055	0.08	0.147	0.159	0.058	0.078	0.038	0.038	0.048	0.056	0.093	0.145	0.147	0.044	0.076	0.077
d6	0.039	0.038	0.046	0.055	0.08	0.147	0.159	0.058	0.078	0.038	0.038	0.048	0.056	0.093	0.145	0.147	0.044	0.076	0.077
d18	0.04	0.039	0.045	0.054	0.067	0.133	0.143	0.061	0.073	0.04	0.04	0.045	0.054	0.068	0.183	0.115	0.044	0.074	0.073
c30	0.041	0.048	0.04	0.057	0.064	0.129	0.147	0.056	0.073	0.04	0.05	0.04	0.059	0.07	0.139	0.12	0.043	0.07	0.071

APPENDIX C

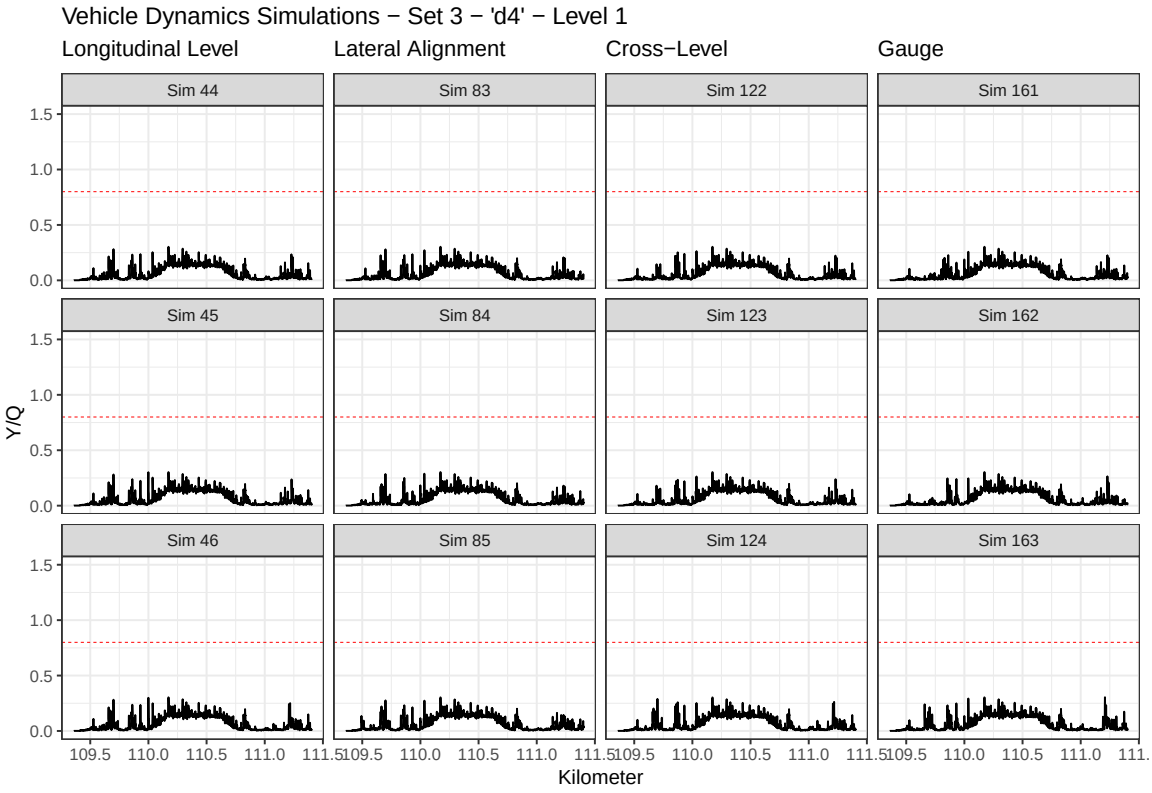

Figure C.1: Phase 3 simulations - wavelet ’d4’ - Level 1

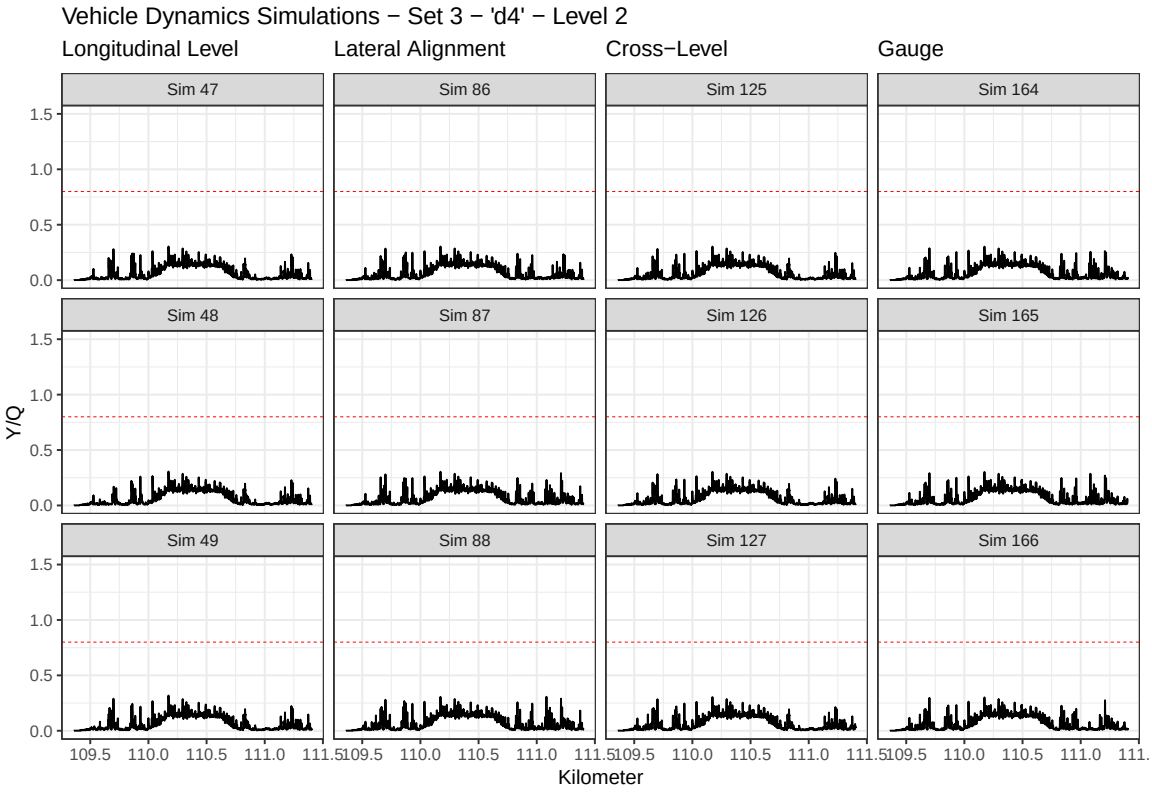

Figure C.2: Phase 3 simulations - wavelet ’d4’ - Level 2

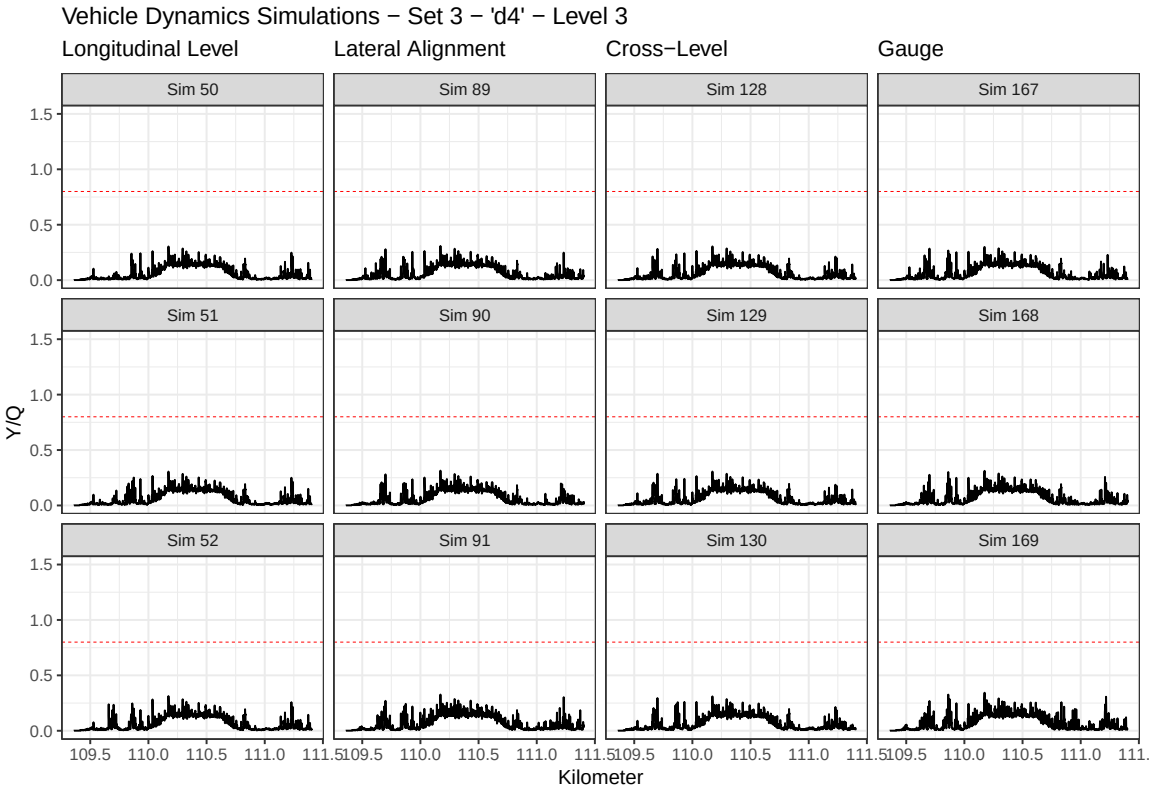

Figure C.3: Phase 3 simulations - wavelet ’d4’ - Level 3

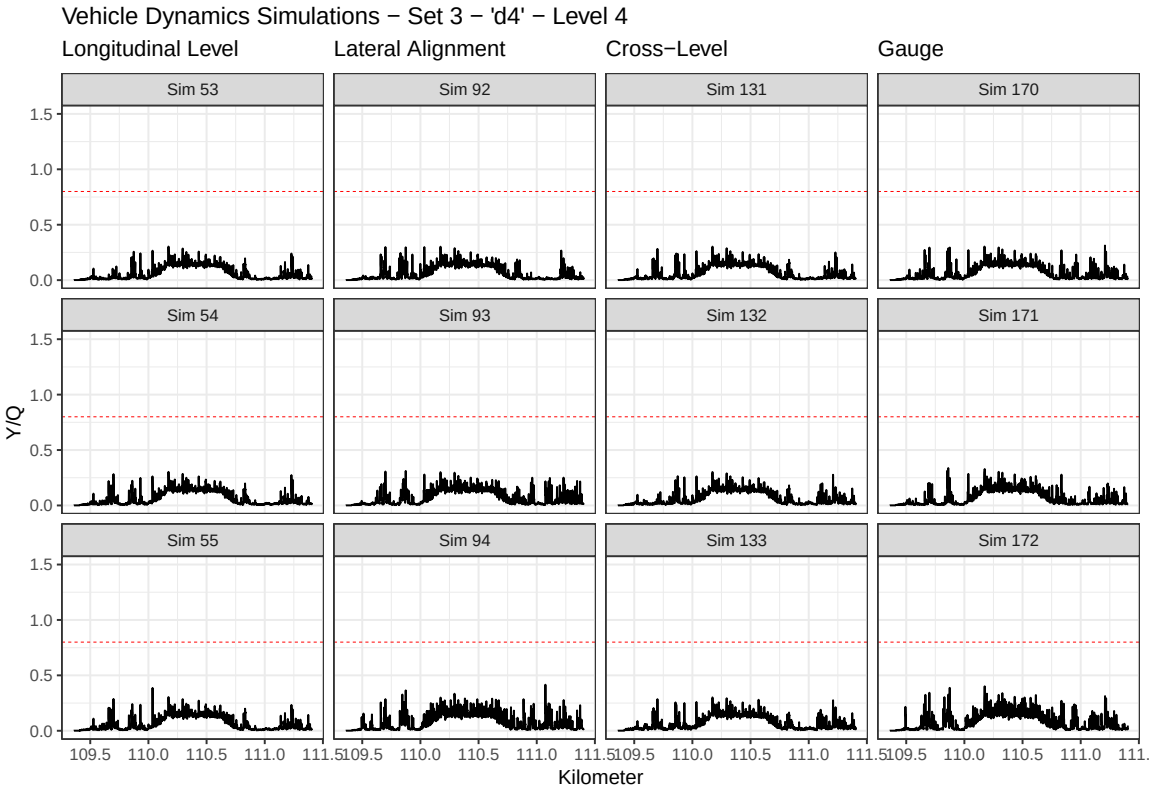

Figure C.4: Phase 3 simulations - wavelet ’d4’ - Level 4

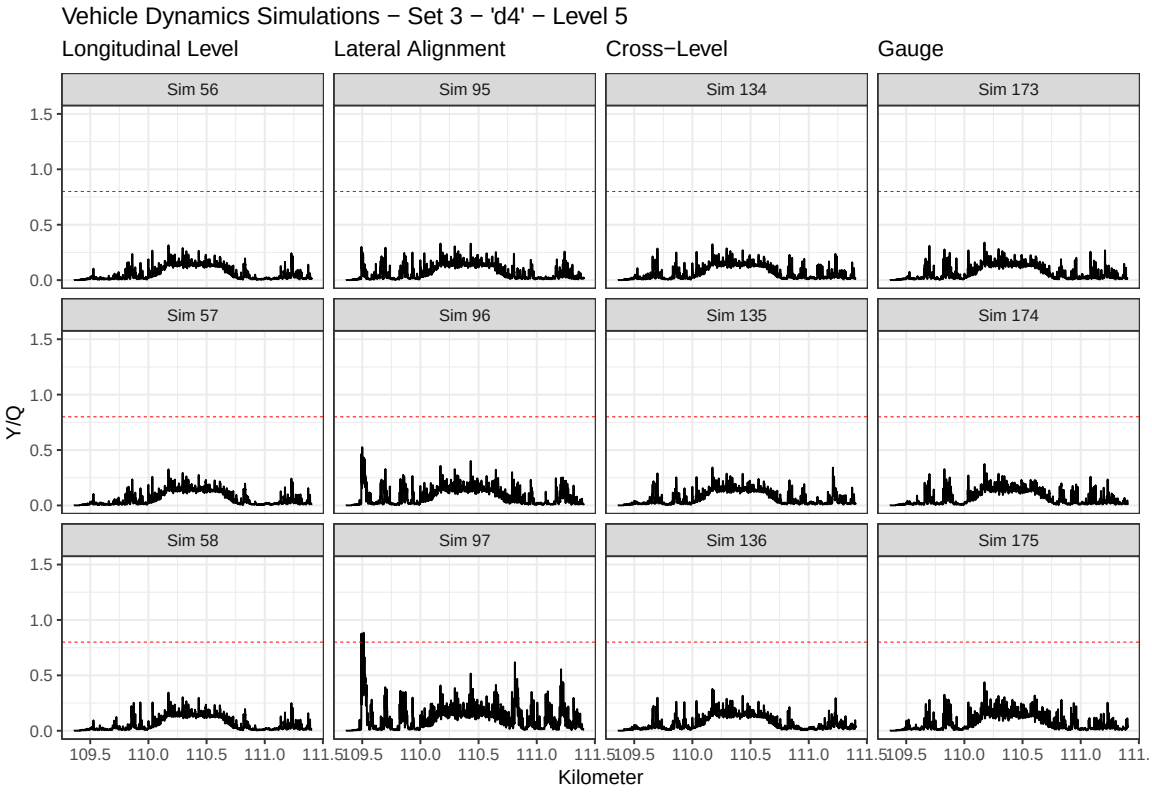

Figure C.5: Phase 3 simulations - wavelet ’d4’ - Level 5

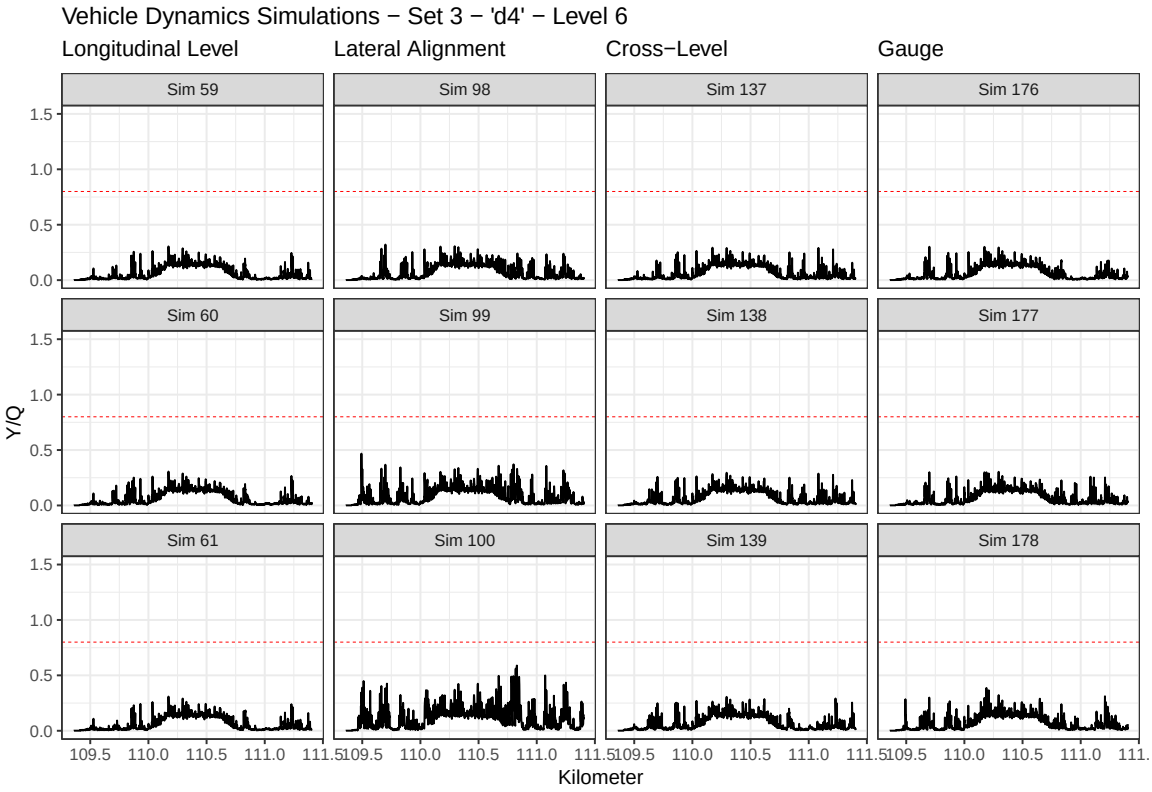

Figure C.6: Phase 3 simulations - wavelet ’d4’ - Level 6

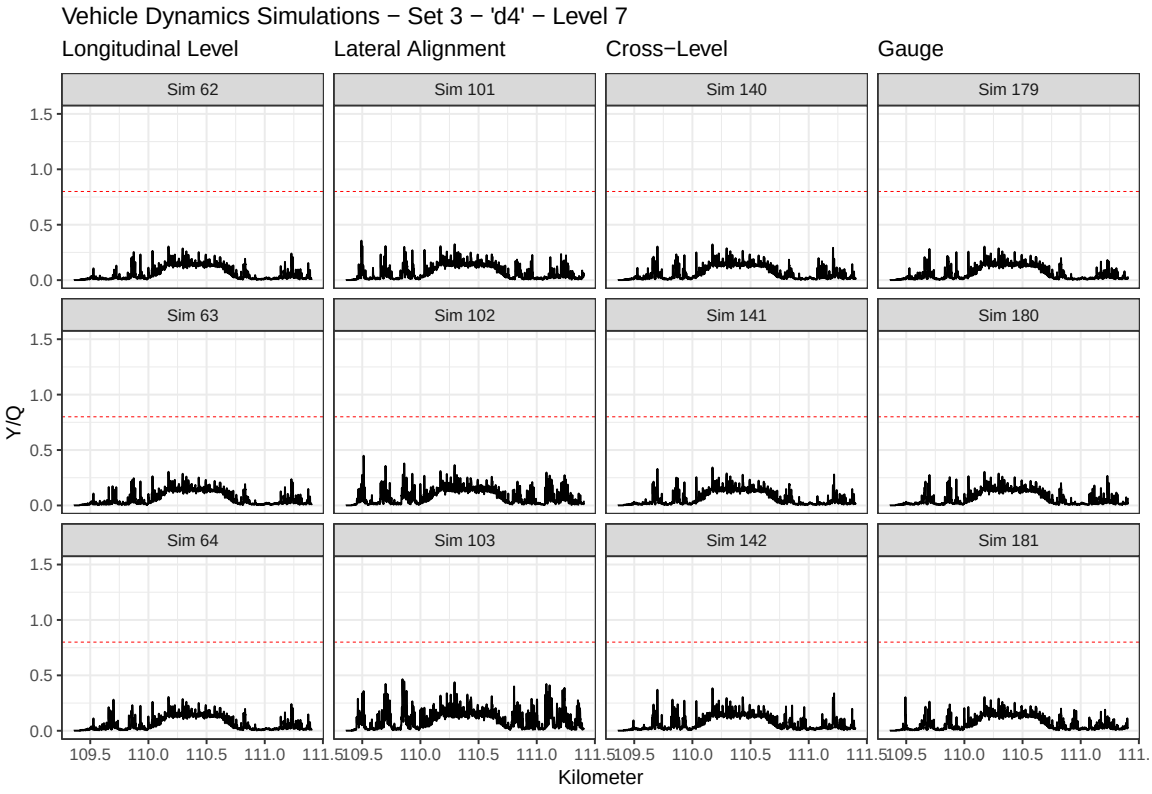

Figure C.7: Phase 3 simulations - wavelet ’d4’ - Level 7

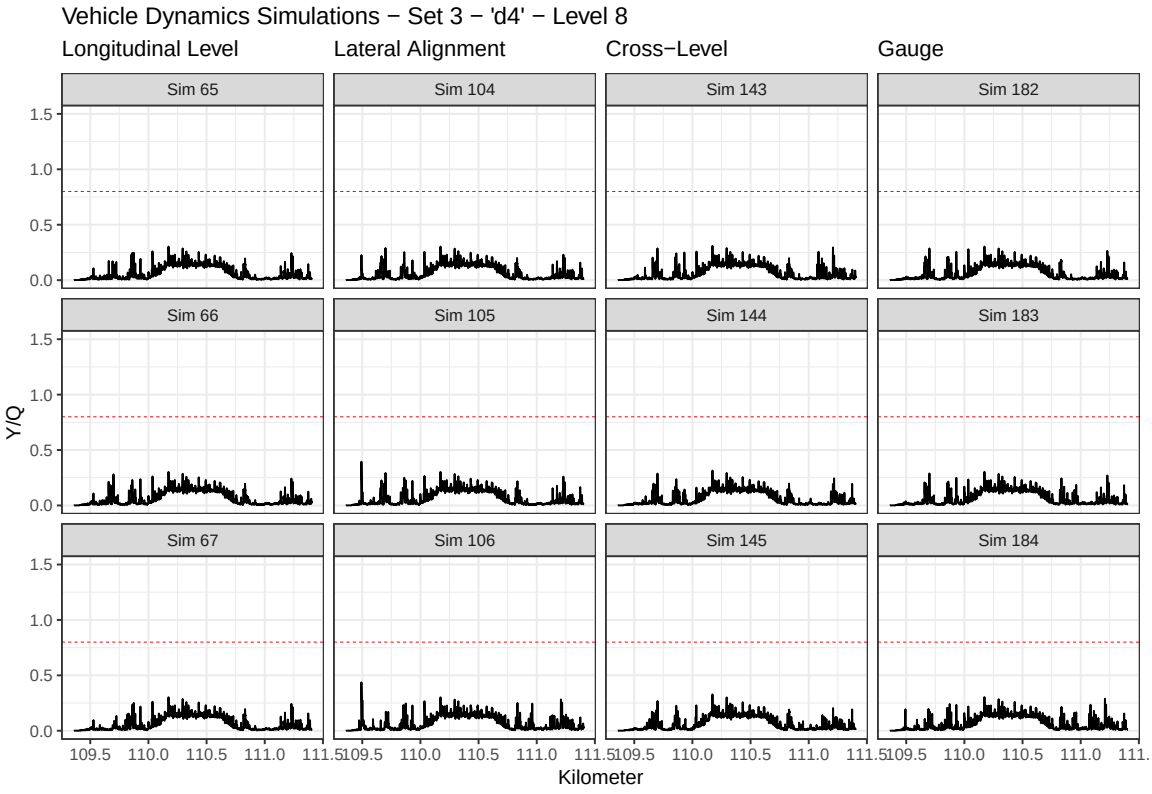

Figure C.8: Phase 3 simulations - wavelet ’d4’ - Level 8

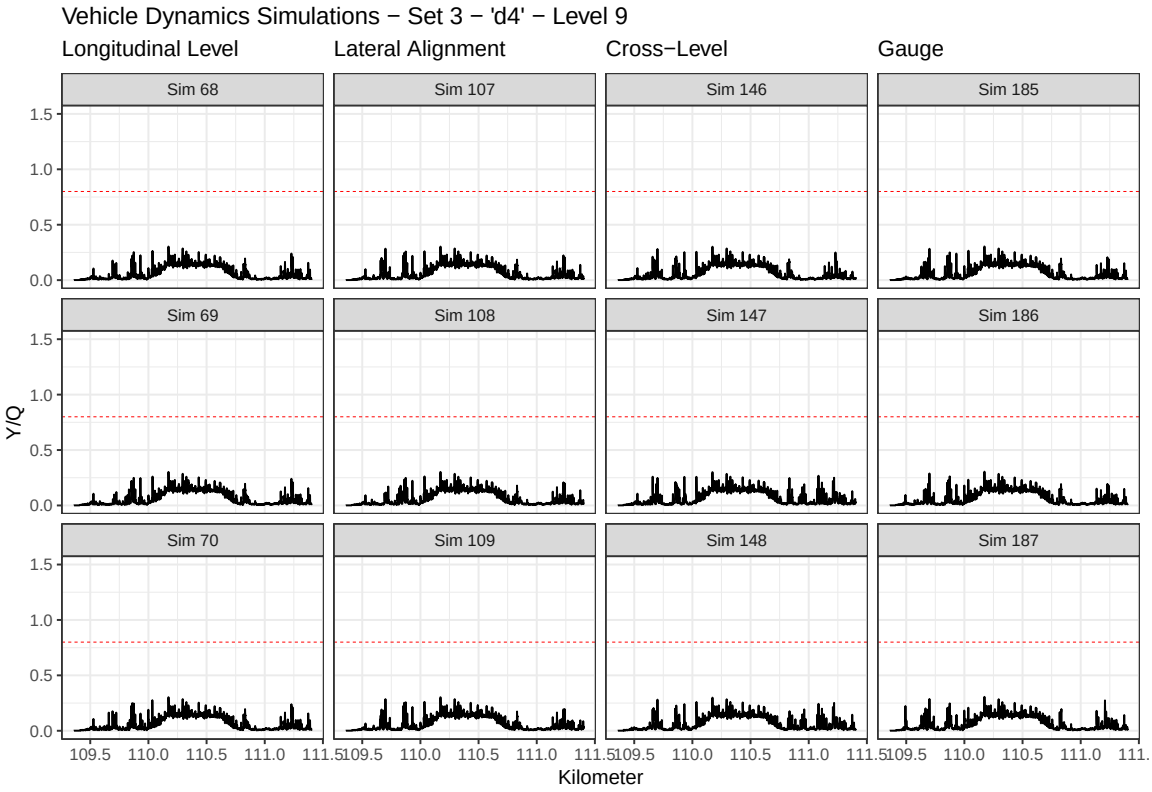

Figure C.9: Phase 3 simulations - wavelet ’d4’ - Level 9

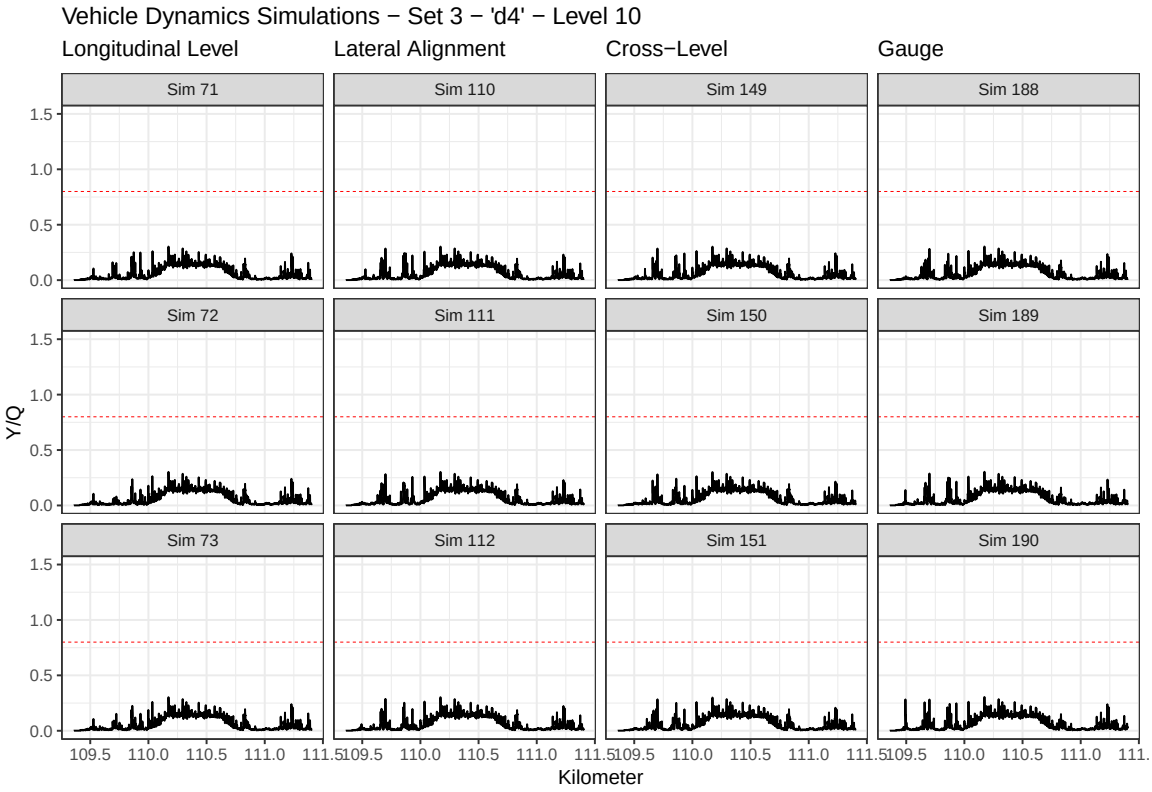

Figure C.10: Phase 3 simulations - wavelet ’d4’ - Level 10

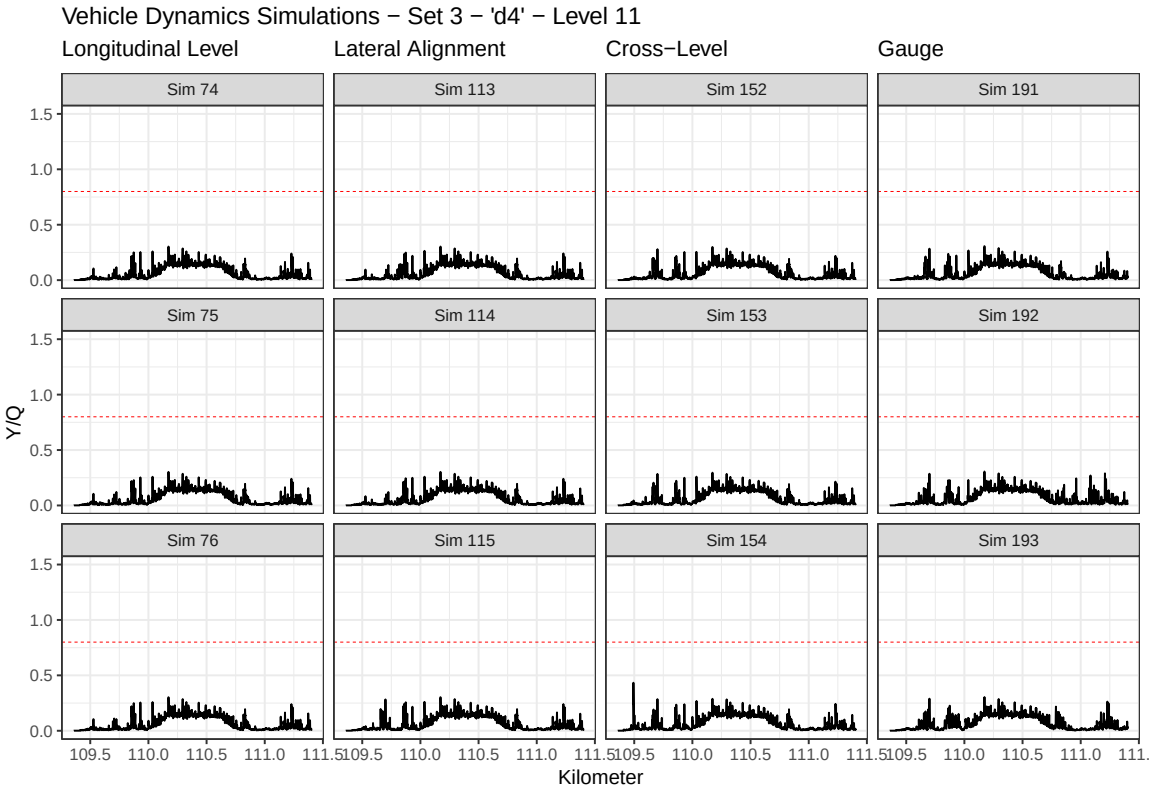

Figure C.11: Phase 3 simulations - wavelet ’d4’ - Level 11

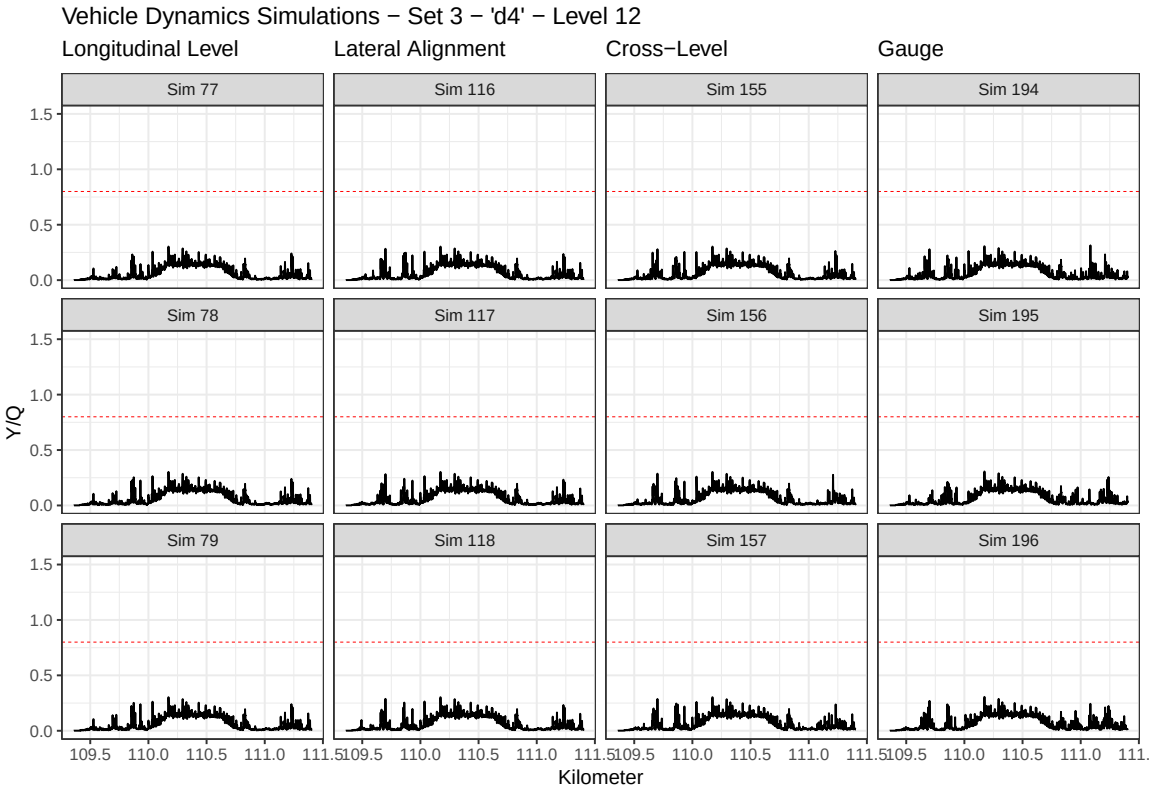

Figure C.12: Phase 3 simulations - wavelet ’d4’ - Level 12

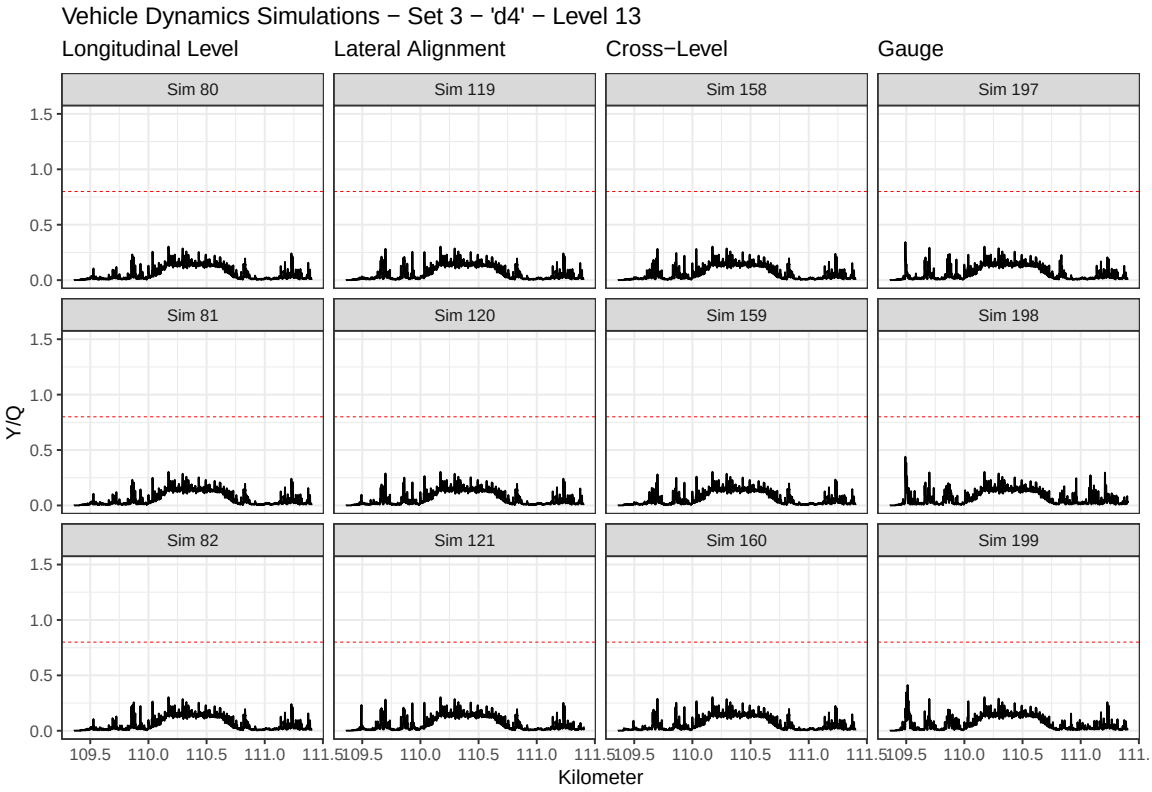

Figure C.13: Phase 3 simulations - wavelet ’d4’ - Level 13

REFERENCES

[1] BSI, *European Standard EN 13848: Railway applications – Track – Track geometry quality - Part 5: Geometry quality levels*, 2008.

[2] UIC, *UIC 714 Leaflet: Classification of lines for the purpose of track maintenance. 4th edition, 2009*, 2009.

[3] J. Lin, M. Asplund, and A. Parida, "Reliability analysis for degradation of locomotive wheels using parametric bayesian approach," *Quality and Reliability Engineering International*, vol. 30, no. 5, pp. 657–667, 2014.

[4] A. Alemi, F. Corman, and G. Lodewijks, "Condition monitoring approaches for the detection of railway wheel defects," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 8, no. 231, pp. 961–981, 2017.

[5] Z. Jiang, D. Banjevic, E. Mingcheng, A. Jardine, and Q. Li, "Remaining useful life estimation of metropolitan train wheels considering measurement error," *Journal of Quality in Maintenance Engineering*, vol. 24, no. 4, pp. 422–436, 2018.

[6] I. Arasteh-khouy, "Cost-effective maintenance of railway track geometry: a shift from safety limits to maintenance limits," Ph.D. dissertation, 2013, p. 144, ISBN: 9789174397024.

[7] E. T. Selig and J. M. Waters, *Track Geotechnology and Substructure Management*. 1994, p. 144.

[8] I. Soleimanmeigouni, A. Ahmadi, and U. Kumar, "Track geometry degradation and maintenance modelling: A review," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 232, no. 1, pp. 73–102, 2018.

[9] C. Esveld, *Modern Railway Track*, 2nd ed. 2001, p. 144.

[10] S. Mohammadzadeh, M. Sangtarashha, and H. Molatefi, "A novel method to estimate derailment probability due to track geometric irregularities using reliability techniques and advanced simulation methods," *Archive of Applied Mechanics*, vol. 81, no. 11, pp. 1621–1637, 2011.

[11] E. G. Berggren, M. X. Li, and J. Spännar, "A new approach to the analysis and presentation of vertical track geometry quality and rail roughness," *Wear*, vol. 265, no. 9-10, pp. 1488–1496, 2008.

[12] UIC, *UIC Leaflet 518: Testing and approval of railway vehicle from the point of view of their dynamic behavior – Safety – Track fatigue – Running behavior. 4th edition, 2009, ISBN 9782746111624.* 2009.

[13] J. Zhao, A. H. C. Chan, and M. P. N. Burrow, "Reliability analysis and maintenance decision for railway sleepers using track condition information," *Journal of the Operational Research Society*, vol. 58, no. 8, pp. 1047–1055, 2007.

[14] A. R. Andrade, "Prediction and optimization of maintenance and renewal actions related to rail track geometry," Ph.D. dissertation, 2014, p. 184.

[15] F. Peng and Y. Ouyang, "Optimal clustering of railroad track maintenance jobs," *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 4, pp. 235–247, 2014.

[16] I. Arasteh-khouy, H. Schunnesson, U. Juntti, A. Nissen, and P. O. Larsson Kråik, "Evaluation of track geometry maintenance for a heavy haul railroad in Sweden: a case study," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 228, no. 5, pp. 496–503, 2014.

[17] T. Zhang, J. Andrews, and R. Wang, "Optimal scheduling of track maintenance on a railway network," *Quality and Reliability Engineering International*, vol. 29, no. 2, pp. 285–297, 2013.

[18] M. A. Costa, J. A. P. Braga, and A. R. Andrade, "A data-driven maintenance policy for railway wheelset based on survival analysis and markov decision process," *Quality and Reliability Engineering International*, pp. 1–23, 2020, doi:10.1002/qre.2729.

[19] ——, "Assessing the performance of different devices in railway wheelset inspection," *Measurement*, vol. 165, pp. 1–14, 2020, doi:10.1016/j.measurement.2020.108145.

[20] M. A. Costa, H. A. Ribeiro, and D. Goldsman, "Spatiotemporal kriging supporting railroad track maintenance decisions," *Proceedings of the International Heavy Haul Association STS Conference, IHHA 2019, June 12–14, Narvik, Norway*, pp. 233–240, 2019.

[21] M. A. Costa, A. R. Andrade, and J. N. Costa, "Statistical representation of railway track irregularities using wavelets," *IAI2020: International Congress and Workshop on Industrial AI 2020*, [forthcoming].

[22] BSI, *BS EN 14363:2016+A1:2018. Railway applications. Testing and Simulation for the acceptance of running characteristics of railway vehicles. Running Behaviour and stationary tests.* 2016.

[23] F. Ghofrani, Q. He, R. M. Goverde, and X. Liu, "Recent applications of big data analytics in railway transportation systems: A survey," *Transportation Research Part C: Emerging Technologies*, vol. 90, no. January, pp. 226–246, 2018.

[24] S. Pradhan, A. K. Samantaray, and R. Bhattacharyya, "Prediction of railway wheel wear and its influence on the vehicle dynamics in a specific operating sector of Indian railways network," *Wear*, vol. 406-407, pp. 92–104, 2018.

[25] G. Tao, Z. Wen, X. Zhao, and X. Jin, "Effects of wheel rail contact modelling on wheel wear simulation," *Wear*, vol. 366-367, pp. 146–156, 2016.

[26] Y. Muhamedsalih, J. Stow, and A. Bevan, "Use of railway wheel wear and damage prediction tools to improve maintenance efficiency through the use of economic tyre turning," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 233, no. 1, pp. 103–117, 2019.

[27] D. Nikas, J. Ahlstrom, and A. Malakizadi, "Mechanical properties and fatigue behaviour of railway wheel steels as influenced by mechanical and thermal loadings," *Wear*, vol. 366-367, pp. 407–415, 2016.

[28] S. C. Wu, Y. X. Liu, C. H. Li, G. Z. Kang, and S. L. Liang, "On the fatigue performance and residual life of intercity railway axles with inside axle boxes," *Engineering Fracture Mechanics*, vol. 197, pp. 176–191, 2018.

[29] A. Shebani and S. Iwnicki, "Prediction of wheel and rail wear under different contact conditions using artificial neural networks," *Wear*, vol. 406-407, no. March 2017, pp. 173–184, 2018.

[30] M. A. Cremona, B. Liu, Y. Hu, S. Bruni, and R. Lewis, "Predicting railway wheel wear under uncertainty of wear coefficient, using universal kriging," *Reliability Engineering and System Safety*, vol. 154, pp. 49–59, 2016.

[31] "Impact analysis due to multiple wheel flats in three dimensional railway vehicle track system model and development of a smart wheelset," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 2, pp. 450–471, 2016.

[32] Y. B. Huang, L. B. Shi, X. J. Zhao, Z. B. Cai, Q. Y. Liu, and W. J. Wang, "On the formation and damage mechanism of rolling contact fatigue surface cracks of wheel/rail under the dry condition," *Wear*, vol. 400-401, pp. 62–73, 2018.

[33] M. Asplund, M. Palo, S. Famurewa, and M. Rantatalo, "A study of railway wheel profile parameters used as indicators of an increased risk of wheel defects," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 2, pp. 323–334, 2016.

[34] G. Krummenacher, C. S. Ong, S. Koller, S. Kobayashi, and J. M. Buhmann, "Wheel Defect Detection with Machine Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 4, pp. 1176–1187, 2018.

[35] G. Zhang and R. Ren, "Study on typical failure forms and causes of high speed railway wheels," *Engineering Failure Analysis*, vol. 105, pp. 1287–1295, 2019.

[36] R. Faddoul, W. Raphael, and A. Chateauneuf, "Maintenance optimization of series systems subject to reliability constraints," *Reliability Engineering and System Safety*, vol. 180, no. November 2017, pp. 179–188, 2018.

[37] Q. Sun, Z. S. Ye, and N. Chen, "Optimal Inspection and Replacement Policies for Multi-Unit Systems Subject to Degradation," *IEEE Transactions on Reliability*, vol. 67, no. 1, pp. 404–413, 2018.

[38] K. T. P. Nguyen, P. Do, K. T. Huynh, C. Bérenguer, and A. Grall, "Joint optimization of monitoring quality and replacement decisions in condition-based maintenance," *Reliability Engineering and System Safety*, vol. 189, pp. 177–195, 2019.

[39] J. S. Ivy and H. B. Nembhard, "A modeling approach to maintenance decisions using statistical quality control and optimization," *Quality and Reliability Engineering International*, vol. 21, no. 4, pp. 355–366, 2005.

[40] P. C. L. Gerum, A. Altay, and M. Baykal-Gürsoy, "Data driven predictive maintenance scheduling policies for railways," *Transportation Research Part C: Emerging Technologies*, vol. 107, pp. 137–154, 2019.

[41] Z. Jiang, D. Banjevic, E. M. Mingcheng, and B. Li, "Optimizing the re-profiling policy regarding metropolitan train wheels based on a semi-Markov decision process," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 231, no. 5, pp. 495–507, 2017.

[42] E. Mingcheng, B. Li, Z. Jiang, and Q. Li, "An optimal reprofiling policy for high-speed train wheels subject to wear and external shocks using a semi-Markov decision process," *IEEE Transactions on Reliability*, vol. 67, no. 4, pp. 1468–1481, 2018.

[43] J. A. Braga and A. R. Andrade, "Optimizing maintenance decisions in railway wheelsets: A Markov decision process approach," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 233, no. 2, pp. 285–300, 2019.

[44] S. M. Ross, *Introduction to Probability Models*, 11th ed. Oxford, UK: Academic Press, 2014.

[45] T. J. Sheskin, *Markov Chains and decision processes for engineers and managers*, 1st ed. Boca Raton, FL: Boca Raton, FL, USA: CRC Press, 2019.

[46] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, New Jersey: New York, NY, USA: John Wiley & Sons Inc., 2005.

[47] D. R. Cox, "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 34, no. 2, pp. 187–220, 1972.

[48] B. Vidakovic, *Engineering Biostatistics: An Introduction using MATLAB and WinBUGS*, 1st ed. Hoboken, NJ: John Wiley & Sons Inc., 2017.

[49] I. Chadés, G. Chapron, M. J. Cros, F. Garcia, and R. Sabbadin, "MDPtoolbox: A multi-platform toolbox to solve stochastic dynamic programming problems," *Ecography*, vol. 37, no. 9, pp. 916–920, 2014.

[50] G. G. Yin and Q. Zhang, *Discrete-Time Markov Chains: two-time-scale methods and applications*, 1st ed. New York, NY, USA: Springer-Verlag, 2005.

[51] D. R. Cox and D. Oakes, *Analysis of Survival Data*. London, UK: Chapman & Hall, 1984.

[52] A. R. Andrade and J. Stow, "Statistical Modelling of Wear and Damage Trajectories of Railway Wheelsets," *Quality and Reliability Engineering International*, vol. 32, no. 8, pp. 2909–2923, 2016.

[53] ——, "Assessing the efficiency of maintenance operators: A case study of turning railway wheelsets on an under-floor wheel lathe," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 231, no. 2, pp. 155–163, 2017.

[54] ——, "Assessing the potential cost savings of introducing the maintenance option of 'Economic Tyre Turning' in Great Britain railway wheelsets," *Reliability Engineering and System Safety*, vol. 168, no. May, pp. 317–325, 2017.

[55] L. Mira, A. R. Andrade, and M. C. Gomes, "Maintenance scheduling within rolling stock planning in railway operations under uncertain maintenance durations," *Journal of Rail Transport Planning and Management*, vol. 14, pp. 1–15, 2020.

[56] S. Madanat and M. Ben-Akiva, "Optimal Inspection and Repair Policies for Infrastructure Facilities," *Transportation Science*, vol. 28, no. 1, pp. 55–62, 1994.

[57] H. C. de Vet, C. B. Terwee, D. L. Knol, and L. M. Bouter, "When to use agreement versus reliability measures," *Journal of Clinical Epidemiology*, vol. 59, no. 10, pp. 1033–1039, 2006.

[58] M. Asplund and J. Lin, "Evaluating the measurement capability of a wheel profile measurement system by using GRR," *Measurement: Journal of the International Measurement Confederation*, vol. 92, pp. 19–27, 2016.

[59] A. Cavuto, M. Martarelli, G. Pandarese, G. Revel, and E. Tomasini, "Train wheel diagnostics by laser ultrasonics," *Measurement: Journal of the International Measurement Confederation*, vol. 80, pp. 99–107, 2016.

[60] R. B. Abernethy, R. P. Benedict, and R. B. Dowdell, "Asme Measurement Uncertainty," *American Society of Mechanical Engineers (Paper)*, vol. 107, no. 83, 1983.

[61] X. X. Yuan and M. D. Pandey, "A nonlinear mixed-effects model for degradation data obtained from in-service inspections," *Reliability Engineering and System Safety*, vol. 94, no. 2, pp. 509–519, 2009.

[62] H. X. Barnhart, M. J. Haber, and L. I. Lin, "An overview on assessing agreement with continuous measurements," *Journal of Biopharmaceutical Statistics*, vol. 17, no. 4, pp. 529–569, 2007.

[63] L. Lin, A. S. Hedayat, B. Sinha, and M. Yang, "Statistical methods in assessing agreement: Models, issues, and tools," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 257–270, 2002.

[64] J. M. Bland and D. G. Altman, "Agreement between methods of measurement with multiple observations per individual," *Journal of Biopharmaceutical Statistics*, vol. 17, no. 4, pp. 571–582, 2007.

[65] H. X. Barnhart, J. Song, and M. J. Haber, "Assessing intra, inter and total agreement with replicated readings," *Statistics in Medicine*, vol. 24, no. 9, pp. 1371–1384, 2005.

[66] P. K. Choudhary and H. N. Nagaraja, "Assessment of agreement using intersection-union principle," *Biometrical Journal*, vol. 47, no. 5, pp. 674–681, 2005.

[67] P. K. Choudhary, "Semiparametric regression for assessing agreement using tolerance bands," *Computational Statistics and Data Analysis*, vol. 51, no. 12, pp. 6229–6241, 2007.

[68] ——, "A tolerance interval approach for assessment of agreement in method comparison studies with repeated measurements," *Journal of Statistical Planning and Inference*, vol. 138, no. 4, pp. 1102–1115, 2008.

[69] C. C. Gluer, G. Blake, Y. Lu, B. A. Blunt, M. Jergas, and H. K. Genant, "Accurate assessment of precision errors: How to measure the reproducibility of bone densitometry techniques," *Osteoporosis International*, vol. 5, no. 4, pp. 262–270, 1995.

[70] M. R. Dann and M. A. Maes, "Stochastic corrosion growth modeling for pipelines using mass inspection data," *Reliability Engineering and System Safety*, vol. 180, pp. 245–254, 2018.

[71] F. Zhang, J. Shen, and Y. Ma, "Optimal maintenance policy considering imperfect repairs and non-constant probabilities of inspection errors," *Reliability Engineering and System Safety*, vol. 193, p. 106 615, 2020.

[72] G. Pulcini, "A perturbed gamma process with statistically dependent measurement errors," *Reliability Engineering and System Safety*, vol. 152, pp. 296–306, 2016.

[73] J. Lin, J. Pulido, and M. Asplund, "Reliability analysis for preventive maintenance based on classical and Bayesian semi-parametric degradation approaches using locomotive wheelsets as a case study," *Reliability Engineering and System Safety*, vol. 134, pp. 143–156, 2015.

[74] P. Urda, S. Munoz, J. F. Aceituno, and J. L. Escalona, "Wheel-rail contact force measurement using strain gauges and distance lasers on a scaled railway vehicle," *Mechanical Systems and Signal Processing*, vol. 138, p. 106 555, 2020.

[75] Y. Huang, C. Huang, J. Ding, and Z. Liu, "Fault diagnosis on railway vehicle bearing based on fast extended singular value decomposition packet," *Measurement: Journal of the International Measurement Confederation*, vol. 152, 2020.

[76] N. Bosso, A. Gugliotta, and N. Zampieri, "Wheel flat detection algorithm for onboard diagnostic," *Measurement: Journal of the International Measurement Confederation*, vol. 123, pp. 193–202, 2018.

[77] A. Alemi, F. Corman, Y. Pang, and G. Lodewijks, "Reconstruction of an informative railway wheel defect signal from wheel-rail contact signals measured by multiple wayside sensors," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 233, no. 1, pp. 49–62, 2019.

[78] ——, "Evaluation of the influential parameters contributing to the reconstruction of railway wheel defect signals," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 233, no. 1, pp. 1–12, 2019.

[79] Z. Chi, J. Lin, R. Chen, and S. Huang, "Data-driven approach to study the polygonization of high-speed railway train wheelsets using field data of China's HSR train," *Measurement: Journal of the International Measurement Confederation*, vol. 149, 2020.

[80] A. Galecki and T. Burzykowski, *Linear Mixed-Effects Models Using R A Step-by-Step Approach*. New York, NY, USA: Springer, 2013.

[81] A. C. Rencher and G. B. Schaalje, *Linear Models in Statistics*. New York, NY, USA: John Wiley & Sons Inc., 2008.

[82] F. Peng, S. Kang, X. Li, Y. Ouyang, K. Somani, and D. Acharya, "A heuristic approach to the railroad track maintenance scheduling problem," *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, pp. 129–145, 2011.

[83] F. Peng and Y. Ouyang, "Track maintenance production team scheduling in railroad networks," *Transportation Research Part B: Methodological*, vol. 46, pp. 1474–1488, 2012.

[84] D. Uzarski and S. McNeil, "Technologies for planning railroad track maintenance and renewal," *Journal of Transportation Engineering*, vol. 120, no. 5, pp. 807–820, 1994.

[85] M. F. Gorman and J. J. Kanet, "Formulation and solution approaches to the rail maintenance production gang scheduling problem," *Journal of Transportation Engineering*, vol. 136, no. 8, pp. 701–708, Aug. 2010.

[86] M. Lake, L. Ferreira, and M. Murray, "Minimising costs in scheduling railway track maintenance," *Computers in Railways VII*, 2000.

[87] J. Pombo and J. Ambrosio, "An alternative method to include track irregularities in railway vehicle dynamic analyses," *Nonlinear Dynamics: An International Journal of Nonlinear Dynamics and Chaos in Engineering Systems*, vol. 68, no. 1-2, pp. 161–176, 2012.

[88] M. X. D. Li, E. G. Berggren, and M. Berg, "Assessment of vertical track geometry quality based on simulations of dynamic track-vehicle interaction," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 223, no. 2, pp. 131–139, 2009.

[89] C. C. Ebeling, *An Introduction to Reliability and Maintainability Engineering*, 3rd ed. Waveland Press, Inc., 2019.

[90] E. Berggren, "Railway track stiffness : Dynamic measurements and evaluation for efficient maintenance," Ph.D. dissertation, KTH, Aeronautical and Vehicle Engineering, 2009.

[91] X. Zhang, N. Feng, Y. Wang, and Y. Shen, "Acoustic emission detection of rail defect based on wavelet transform and shannon entropy," *Journal of Sound and Vibration*, vol. 339, pp. 419–432, 2015.

[92] L. M. Quiroga and E. Schnieder, "A heuristic approach to railway track maintenance scheduling," *WIT Transactions on the Built Environment*, vol. 114, pp. 687–699, 2010.

[93] T. Oyama and M. Miwa, "Mathematical modeling analyses for obtaining an optimal railway track maintenance schedule," *Japan Journal of Industrial and Applied Mathematics*, vol. 23, pp. 207–224, 2006.

[94] I. Arasteh-khouy, U. Juntti, A. Nissen, and H. Schunnesson, "Evaluation of track geometry degradation in swedish heavy haul railroad - a case study," *International Journal of COMADEM*, vol. 15, pp. 11–16, 2012.

[95] I. Arasteh-khouy, H. Schunnesson, U. Juntti, A. Nissen, and P.-O. Larsson-Kråik, "Evaluation of track geometry maintenance for a heavy haul railroad in sweden - a case study," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, pp. 1–8, 2013.

[96] I. Arasteh-khouy, P.-O. Larsson-Kråik, A. Nissen, J. Lundberg, and U. Kumar, "Geometrical degradation of railway turnouts - a case study from a swedish heavy haul railroad," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 2013.

[97] A. Higgins, "Scheduling of railway track maintenance activities and crews," *Journal of the Operational Research Society*, 1998.

[98] A. R. Andrade and P. F. Teixeira, "Biobjective optimization model for maintenance and renewal decisions related to rail track geometry," *Transportation research record*, vol. 2261, no. 1, pp. 163–170, 2011.

[99] C. Vale, I. M. Ribeiro, and R. Calçada, "Integer programming to optimize tamping in railway tracks as preventive maintenance," *Journal of Transportation Engineering*, vol. 138, no. 1, pp. 123–131, 2012.

[100] M. Shafiee, M. Patriksson, and S. Chukova, "An optimal age–usage maintenance strategy containing a failure penalty for application to railway tracks," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 2, pp. 407–417, 2016.

[101] A. R. Andrade and P. F. Teixeira, "A bayesian model to assess rail track geometry degradation through its life-cycle," *Research in transportation economics*, vol. 36, no. 1, pp. 1–8, 2012.

[102] ——, "Hierarchical bayesian modelling of rail track geometry degradation," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit*, vol. 227, no. 4, pp. 364–375, 2013.

[103] A. R. Andrade and P. F. Teixeira, "Statistical modelling of railway track geometry degradation using hierarchical bayesian models," *Reliability Engineering & System Safety*, vol. 142, pp. 169–183, 2015.

[104] A. R. Andrade and P. F. Teixeira, "Exploring different alert limit strategies in the maintenance of railway track geometry," *Journal of Transportation Engineering*, vol. 142, no. 9, p. 04 016 037, 2016.

[105] L. Bai, R. Liu, Q. Sun, F. Wang, and F. Wang, "Classification-learning-based framework for predicting railway track irregularities," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 2, pp. 598–610, 2016.

[106] B. Bergquist and P. Söderholm, *Improved Condition Assessment through Statistical Analyses: Case Study of Railway Track*. Luleå University of Technology, 2017.

[107] M. Shenton, "Ballast deformation and track deterioration," *Track technology*, pp. 253–265, 1985.

[108] J. Öberg and E. Andersson, "Determining the deterioration cost for railway tracks," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 223, no. 2, pp. 121–129, 2009.

[109] T. Dahlberg, "Some railroad settlement models—a critical review," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 215, no. 4, pp. 289–300, 2001.

[110] L. M. Quiroga and E. Schnieder, "Monte carlo simulation of railway track geometry deterioration and restoration," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 226, no. 3, pp. 274–282, 2012.

[111] T. J. Santner, W. J. Brian, and W. I. Notz, *The Design and Analysis of Computer Experiments*. New York, NY, USA: Springer, 2003.

[112] V. R. Joseph, "Limit kriging," *Technometrics*, vol. 48, no. 4, pp. 458–466, 2006.

[113] B. Bergquist and P. Soderholm, "Data analysis for condition-based railway infrastructure maintenance," *Quality and Reliability Engineering International*, vol. 31, no. 5, pp. 773–781, 2015.

[114] Y. Zhang and W. I. Notz, "Computer experiments with qualitative and quantitative variables: A review and reexamination," *Quality Engineering*, vol. 27, no. 1, pp. 2–13, 2015.

[115] Y. Hung, V. R. Joseph, and S. N. Melkote, "Analysis of computer experiments with functional response," *Technometrics*, vol. 57, no. 1, pp. 35–44, 2015.

[116] L. N. VanBrackle and M. R. Reynolds, "EWMA and CUSUM control charts in the presence of correlation," *Communications in Statistics Part B: Simulation and Computation*, vol. 26, no. 3, pp. 979–1008, 1997.

[117] P. Salvador, V. Naranjo, R. Insa, and P. Teixeira, "Axlebox accelerations: Their acquisition and time-frequency characterisation for railway track monitoring purposes," *Measurement*, vol. 82, pp. 301–312, 2016.

[118] F. L. d. Scalea and J. McNamara, "Wavelet transform for characterizing longitudinal and lateral transient vibrations of railroad tracks," *Research in Nondestructive Evaluation*, vol. 15, no. 2, pp. 87–98, 2004.

[119] F. Balouchi, A. Bevan, and R. Formston, "Development of railway track condition monitoring from multi-train in-service vehicles," *Vehicle System Dynamics*, pp. 1–21, 2020.

[120] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 1483–1510, 2006.

[121] P. Koziol, "Experimental validation of wavelet based solution for dynamic response of railway track subjected to a moving train," *Mechanical Systems and Signal Processing*, vol. 79, pp. 174–181, 2016.

[122] Z. Zhiping, L. Fei, and Z. Yong, "Wavelet analysis of track profile irregularity for beijing-tianjin intercity high speed railway on bridge," in *2010 International Conference on Intelligent Computation Technology and Automation*, IEEE, vol. 3, 2010, pp. 1155–1158.

[123] L. Xu, W. Zhai, and Z. Chen, "On use of characteristic wavelengths of track irregularities to predict track portions with deteriorated wheel/rail forces," *Mechanical Systems and Signal Processing*, vol. 104, pp. 264–278, 2018.

[124] A. Caprioli, A. Cigada, and D. Raveglia, "Rail inspection in track maintenance: A benchmark between the wavelet approach and the more conventional Fourier analysis," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 631–652, 2007.

[125] H. A. Toliyat, S. Member, K. Abbaszadeh, M. M. Rahimian, and L. E. Olson, "Packet Decomposition," *IEEE Transactions of Industry Applications*, vol. 39, no. 5, pp. 1454–1461, 2003.

[126] A. A. Shah, B. S. Chowdhry, T. D. Memon, I. H. Kalwar, and J. A. Ware, "Real time identification of railway track surface faults using canny edge detector and 2d discrete wavelet transform," *Annals of Emerging Technologies in Computing (AETiC)*, vol. 4, no. 2, pp. 53–60, 2020.

[127] B. M. Hopkins and S. Taheri, "Track health monitoring using wavelets," in *Rail Transportation Division Conference*, vol. 44069, 2010, pp. 9–15.

[128] N. Ahuja, S. Lertrattanapanich, and N. Bose, "Properties determining choice of mother wavelet," *IEE Proceedings - Vision, Image and Signal Processing*, vol. 152, no. 5, pp. 659–664, 2005.

[129] A. R. Andrade, T. Johnson, and J. Stow, "Application of wavestrapping statistical technique to estimate an extreme value in train aerodynamics," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 175, pp. 419–427, 2018.

[130] D. B. Percival, S. Sardy, and A. C. Davison, "Wavestrapping time series: Adaptive wavelet-based bootstrapping," *Nonlinear and Nonstationary Signal Processing*, pp. 1–29, 2000.

[131] T. Kojima, H. Tsunashima, and A. Matsumoto, "Fault detection of railway track by multi-resolution analysis," *WIT Transactions on The Built Environment*, vol. 88, 2006.

[132] M. J. Gomez, C. Castejon, E. Corral, and J. C. Garcia-Prada, "Railway axle condition monitoring technique based on wavelet packet transform features and support vector machines," *Sensors*, vol. 20, no. 12, p. 3575, 2020.

[133] C. Higgins and X. Liu, "Modeling of track geometry degradation and decisions on safety and maintenance: A literature review and possible future research directions," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 232, no. 5, pp. 1385–1397, 2018.

[134] B. Vidakovic, *Statistical Modeling by Wavelets*, 1st ed. New York, NY, USA: John Wiley & Sons Inc., 1999.

[135] W. K. Ngui, M. S. Leong, L. M. Hee, and A. M. Abdelrhman, "Wavelet analysis: Mother wavelet selection methods," *Applied Mechanics and Materials*, vol. 393, pp. 953–958, 2013.

[136] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[137] B. N. Singh and A. K. Tiwari, "Optimal selection of wavelet basis function applied to ECG signal denoising," *Digital Signal Processing: A Review Journal*, vol. 16, no. 3, pp. 275–287, 2006.

[138] M. Ahadi and M. S. Bakhtiar, "Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to Acoustic Emission signals," *Applied Acoustics*, vol. 71, no. 7, pp. 634–639, 2010.

[139] H. He, Y. Tan, and Y. Wang, "Optimal base wavelet selection for ecg noise reduction using a comprehensive entropy criterion," *Entropy*, vol. 17, no. 9, pp. 6093–6109, 2015.

[140] L. Yang, M. Judd, and C. J. Bennoch, "Denoising uhf signal for pd detection in transformers based on wavelet technique," Nov. 2004, pp. 166–169, ISBN: 0-7803-8584-5.

[141] P. P. C. Tsui and O. A. Basir, "Wavelet basis selection and feature extraction for shift invariant ultrasound foreign body classification," *IEEE Transactions of Industry Applications*, vol. 45, no. 1–4, pp. 1–14, 2006.

[142] J. Rafiee and P. W. Tse, "Use of autocorrelation of wavelet coefficients for fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 23, no. 5, pp. 1554–1572, 2009.

[143] H. M. Oliveira, "Shannon and renyi entropy of wavelets," *International Journal of Mathematics and Computer Science*, vol. 10, no. 1, pp. 13–26, 2015.

[144] J. Nadal, *Locomotives a Vapeur: Collection Encyclopedie Scientifique*. Paris, FR: Biblioteque de Mecanique Appliquee et Genie, 1908.

[145] BSI, *European Standard EN 14363-1: Railway applications - Testing and Simulation for the acceptance of running characteristics of railway vehicles - Running Behaviour and stationary tests*, 2005.

VITA

Mariana de Almeida Costa received her B.Sc. in Industrial Engineering in 2013 from the Federal University of Juiz de Fora in Brazil. She also completed an MBA in Finance in the same institution in 2014. From 2013-2015 she worked for one of the top railway companies in Brazil, MRS Logística S.A., where she held different positions, first as a trainee, then supply chain analyst, responsible for the procurement and expedition of maintenance parts and, later, railway specialist, engaging in several activities related to the maintenance decision-making process. During 2015-2017 she earned a M.Sc. in Operations Research from the Georgia Institute of Technology. She is currently a Ph.D. candidate in Industrial Engineering, in the Statistics track. Her doctorate research has focused on the development of data-driven statistical models for the analysis and optimization of the maintenance decision-making process in railway track and wheelset maintenance, combining the study of trade-offs between minimum cost solutions and ease of implementation of the recommended policy. During her studies, she had the opportunity to join the European research project SMaRTE (Smart Maintenance and the Rail Traveller Experience) at IDMEC - Instituto Superior Técnico (Portugal) for a 9-month period. This work resulted in an extensive report on the improvement of maintenance strategies for passenger train operating companies in Europe.